



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 114883

TO: Jeffrey Parkin
Location: rem/3e39/3c18
Art Unit: 1648
February 27, 2004

Case Serial Number: 09/891609

From: P. Sheppard
Location: Remsen Building
Phone: (571) 272-2529

sheppard@uspto.gov

Search Notes

This Page Blank (uspto)

STIC-Biotech/ChemLib

114 883

From: Parkin, Jeffrey
Sent: Sunday, February 22, 2004 4:51 PM
To: STIC-Biotech/ChemLib
Subject: U.S. Serial No. 09/891,609

Please search **SEQ ID NOS.: 2 and 4** from the aforementioned application (**09/891,609; Stamatatos et al.**) v. all relevant databases, including interference. Please provide a paper copy and electronic copy of the results.
Thanks!

JSP
AU 1648
REM 3E39
2-0908

RECEIVED
FEB 23 2004
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:02:44 ; Search time 55.1526 Seconds
(without alignments)
3309.468 Million cell updates/sec

Title: US-09-891-609A-2
Perfect score: 3474
Sequence: 1 MRVKGIRKNYQHLWRGGTLL.....DKWASLWNWFDISKWLWYIK 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3474	100.0	646	5	AAU75155 Modified
2	3422	98.5	847	3	AAU75155 Variant H
3	3327	95.8	619	5	AAU75156 N-termina
4	3279.5	94.4	842	6	ABR55684 HIV isola
5	3279.5	94.4	842	7	ADC13218 Protein o
6	3274.5	94.3	842	6	ABU66565 Human imm
7	3264.5	94.0	842	5	ABB06211 HIV Env i
8	3065	88.2	883	4	AAU75156 Ancestral
9	3062.5	88.2	850	2	AAU75156 gpl20 fro
10	3006.5	86.5	856	6	ABR55495 Amino aci
11	2997	86.3	851	1	AAU75156 HIV prote
12	2984	85.9	855	2	AAU75156 ENV prote
13	2984	85.9	855	3	AAU75156 HIV-1 (AT
14	2984	85.9	855	3	AAU75156 HIV-1 (AT
15	2984	85.9	855	6	ABU57553 AIDS asso
16	2984	85.9	855	6	ABU57550 AIDS asso
17	2984	85.9	855	6	ABU63182 Protein #
18	2984	85.9	855	6	ABU63186 Env prote
19	2984	85.9	863	1	AAU75156 Sequence
20	2984	85.9	863	2	AAU75156 env gene
21	2984	85.9	863	5	AAE35790 ARV-2 (9B
22	2978.5	85.7	855	2	AAU75156 Human Imm
23	2978.5	85.7	855	2	AAU75156 Env prote
24	2974	85.6	857	2	AAU75156 gpl20 fro
25	2951.5	85.0	860	2	AAU75156 HIV-SF2 v

26	2940.5	84.6	855	2	AAW43069	Aaw43069 HIV-1 gpl
27	2939.5	84.6	856	2	AAU75156	Aar67726 gpl20 fro
28	2939.5	84.6	856	4	AAU75156	Aab85999 Amino aci
29	2939.5	84.6	863	2	AAU75156	Aar43869 HTLV-III
30	2933	84.4	868	5	AAU75156	Aao19389 Lymphaden
31	2931.5	84.4	856	2	AAU75156	Aar25940 Modified
32	2931.5	84.4	866	1	AAU75156	Aap80966 HIV prote
33	2930	84.3	868	1	AAU75156	Aap60422 Sequence
34	2929.5	84.3	863	1	AAU75156	Aap60349 HTLV-III
35	2926.5	84.2	856	1	AAU75156	Aap61514 Sequence
36	2926.5	84.2	856	2	AAU75156	Aaw89325 HIV-1 env
37	2926.5	84.2	856	6	ABU63322	Abu63322 Human lym
38	2926	84.2	868	1	AAU75156	Aap60063 HIV virus
39	2925	84.2	865	1	AAU75156	Aap70175 Sequence
40	2922.5	84.1	856	3	AAU75156	Aay97072 Wild type
41	2922.5	84.1	856	8	AAU75156	Ade84721 Human imm
42	2919.5	84.0	854	3	AAU75156	Aab10053 HIV-1 coa
43	2919.5	84.0	854	3	AAU75156	Aab10697 HIV-1 env
44	2919.5	84.0	854	4	AAU75156	Aab86199 HIV gp41
45	2919.5	84.0	854	5	ABG73663	ABG73663 HIV-1 NL4

ALIGNMENTS

RESULT 1
AAU75155
ID AAU75155 standard; protein; 646 AA.
XX
AC AAU75155;
XX
DT 08-MAY-2002 (first entry)
XX
DE Modified full-length HIV-1 SF162DV2 gpl40 envelope protein.
XX
KW Human immunodeficiency virus type 1; HIV-1; gpl40 envelope protein;
KW second hypervariable region; V2; SF162deltav2; SF162DV2; AIDS;
KW acquired immunodeficiency syndrome; anti-HIV; mutant; mutein; immunogen.
XX
OS Human immunodeficiency virus 1; strain SF162 (clade B).
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 539
FT /note= "Encoded by CAGCTC"
XX
PN WO200200250-A2.
XX
PD 03-JAN-2002.
XX
PF 27-JUN-2001; 2001WO-US020483.
XX
PR 27-JUN-2000; 2000US-0214608P.
PR 26-JUN-2001; 2001US-00891609.
XX
PA (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI Stamatos L, Barnett S, Shrivastava I;
XX
DR WPI; 2002-130836/17.
DR N-PSDB; AAS15498.
XX
PT Immunizing an animal and eliciting an immune response against
PT heterologous HIV-1 in an animal, involves administering an immunogen
PT comprising modified HIV-1 envelope protein, or DNA or virus encoding the
PT protein.
XX
PS Claim 5; Fig 16; 62pp; English.
XX
CC The present invention relates to methods for immunising an animal,
CC preferably human, against heterologous human immunodeficiency virus type
CC 1 (HIV-1), and eliciting a heterologous immune response to HIV-1 in that
CC animal. The method comprises administering an immunogen having at least

CC one modified HIV-1 envelope protein with a deletion in the second
CC hypervariable (V2) region or its fragment, or DNA or virus encoding the
CC modified HIV-1 envelope protein. The invention discloses the
CC polynucleotide and amino acid sequences for the full-length HIV-1
CC SF162deltav2 (SF162DV2) gp140 envelope protein, and, for an N-terminal
CC mutant of SF162DV2. The immunogens of the invention are useful for
CC immunising an animal against heterologous HIV-1 strains by eliciting
CC neutralising antibodies or protective antibodies in the animal. The
CC method can be used to treat acquired immunodeficiency syndrome (AIDS).
CC The present sequence represents the full-length modified HIV-1 SF162DV2
CC gp140 envelope protein
XX
SQ Sequence 646 AA;

Query Match 100.0%; Score 3474; DB 5; Length 646;
Best Local Similarity 100.0%; Pred. No. 1.9e-169;
Matches 646; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRVKGIRKQNYQHLWRGGTLLGLMLMICS AVEKLVWTVVYGYGVPVWKEATTTLFCASDAKAY 60
Db |||||
1 MRVKGIRKQNYQHLWRGGTLLGLMLMICS AVEKLVWTVVYGYGVPVWKEATTTLFCASDAKAY 60
Qy 61 DTEVHNWVWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120
Db |||||
61 DTEVHNWVWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120
Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAGKGLINCNTSVITQACPKVS 180
Db |||||
121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAGKGLINCNTSVITQACPKVS 180
Qy 181 FEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 240
Db |||||
181 FEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 240
Qy 241 VVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYPATGDIIGDIRQAH 300
Db |||||
241 VVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYPATGDIIGDIRQAH 300
Qy 301 CNISGEKWNNTLKQIVTKLQAOFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 360
Db |||||
301 CNISGEKWNNTLKQIVTKLQAOFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 360
Qy 361 NSTWNNTIGNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 420
Db |||||
361 NSTWNNTIGNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 420
Qy 421 GGKEISNTTEIFRPGGGDMRDNRWRS ELYKYKVVKIEPLGVAPTAKKRVRVQREKRAVTLG 480
Db |||||
421 GGKEISNTTEIFRPGGGDMRDNRWRS ELYKYKVVKIEPLGVAPTAKKRVRVQREKRAVTLG 480
Qy 481 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKLQ 540
Db |||||
481 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKLQ 540
Qy 541 ARVLAVERYLKDQQLLGWCGSKLIC TTAVPWNASWSNKS LDDQIWNNTWMEWEREIDN 600
Db |||||
541 ARVLAVERYLKDQQLLGWCGSKLIC TTAVPWNASWSNKS LDDQIWNNTWMEWEREIDN 600
Qy 601 YTNLIYTLIEESQNQQEKNEQELLELDK WASLNWFDISKWLWYIK 646
Db |||||
601 YTNLIYTLIEESQNQQEKNEQELLELDK WASLNWFDISKWLWYIK 646

RESULT 2
AA97073
ID AA97073 standard; protein; 847 AA.

XX
AC AA97073;
XX
DT 12-SEP-2003 (revised)
DT 31-OCT-2000 (first entry)
XX
DE Variant HIV-1 SF162 Env gp160.

XX HIV-1; vaccine; anti-HIV; immunogenic; modified; envelope; Env;
KW CD4 binding region; V1/V2 loop; bridging sheet.
KW
XX Human immunodeficiency virus 1; isolate SF162.
OS
XX WO200039303-A2.
PN
XX 06-JUL-2000.
PD
XX 30-DEC-1999; 99WO-US031272.
PF
XX 31-DEC-1998; 98US-0114495P.
PR 29-SEP-1999; 99US-0156670P.
XX (CHIR) CHIRON CORP.
PA
XX Barnett S, Hartog K, Martin E;
PI
XX WPI; 2000-465745/40.
DR
XX
PT Novel modified HIV Env polypeptides useful as immunizing agents and for
PT preparing a vaccine to elicit an immune response against a broad range of
PT HIV subtypes.
XX
PS Claim 5; Page 115-117; 139pp; English.
XX
CC Novel immunogenic modified human immunodeficiency virus (HIV) envelope
CC (Env) polypeptides having an amino acid deleted or replaced in the region
CC corresponding to residues 420-436 or 119-123 and 199-210 relative to
CC isolate HXB-2 are disclosed. The modified Env polypeptide is based on HIV
CC strain SF162, with numbering relative to isolate HXB-2. The Env
CC polypeptides are modified so as to expose at least part of the CD4
CC binding region. The modified HIV Env polypeptides, coding polynucleotides
CC and constructs, further comprising an adjuvant, are used for inducing an
CC immune response in an individual. The method involves administering a
CC first composition comprising a polynucleotide encoding the Env
CC polypeptide in a priming step and administering a second composition
CC comprising a modified Env polypeptide as a booster in an amount
CC sufficient to induce an immune response in the individual. The first
CC and/or second composition further comprises an adjuvant (claimed). The
CC intracellularly produced Env polypeptides can be used for a number of
CC diagnostic and therapeutic purposes to determine the presence of reactive
CC antibodies/and or Env proteins in a biological sample to aid in the
CC diagnosis of HIV infection or disease status or as measure of response to
CC immunization. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 847 AA;

Query Match 98.5%; Score 3422; DB 3; Length 847;
Best Local Similarity 95.4%; Pred. No. 1.1e-166;
Matches 643; Conservative 0; Mismatches 3; Indels 28; Gaps 2;

Qy 1 MRVKGIRKQNYQHLWRGGTLLGLMLMICS AVEKLVWTVVYGYGVPVWKEATTTLFCASDAKAY 60
Db |||||
1 MRVKGIRKQNYQHLWRGGTLLGLMLMICS AVEKLVWTVVYGYGVPVWKEATTTLFCASDAKAY 60
Qy 61 DTEVHNWVWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120
Db |||||
61 DTEVHNWVWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120
Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db |||||
121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVTTSIRNKMQKEYALFYKLDVV 180
Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNV 213
Db |||||
181 PIDNDNTSYKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNV 240
Qy 214 STVQCTHGIRPVVSTQLLNGSLAEEGWVIRSENFDTNAKTIIVQLKESVEINCTRPNNN 273
Db |||||
241 STVQCTHGIRPVVSTQLLNGSLAEEGWVIRSENFDTNAKTIIVQLKESVEINCTRPNNN 300

QY 274 TRKSITIGPRAFYATGDIIGDIRQAHCHNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS 333
Db 301 TRKSITIGPRAFYATGDIIGDIRQAHCHNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS 360
QY 334 SGGDPEIVMHSFNCGGEFFYCNSSTQLENSTWNNTIGPNTNGTITLPCRIKQIINRWQEV 393
Db 361 SGGDPEIVMHSFNCGGEFFYCNSSTQLENSTWNNTIGPNTNGTITLPCRIKQIINRWQEV 420
QY 394 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKV 453
Db 421 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKV 480
QY 454 KIEPLGVAPTKAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSILTLTVQARQLLSGIV 513
Db 481 KIEPLGVAPTKAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSILTLTVQARQLLSGIV 540
QY 514 QQONNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVP 572
Db 541 QQONNLLRAIEAQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWCSGKLICTTAVP 600
QY 573 WNASWSNKSLDQIWNNTMTMEWEREIDNYTNLIYTLIEESQKQEKNEQELLELDKWASL 632
Db 601 WNASWSNKSLDQIWNNTMTMEWEREIDNYTNLIYTLIEESQKQEKNEQELLELDKWASL 660
QY 633 WNWFDISKWLWYIK 646
Db 661 WNWFDISKWLWYIK 674

RESULT 3
AAU75156
ID AAU75156 standard; protein; 619 AA.
XX
AC AAU75156;
XX
DT 08-MAY-2002 (first entry)
XX
DE N-terminal mutant of modified HIV-1 SF162DV2 gp140 envelope protein.
XX
KW Human immunodeficiency virus type 1; HIV-1; gp140 envelope protein;
KW second hypervariable region; V2; SF162deltaV2; SF162DV2; AIDS;
KW acquired immunodeficiency syndrome; anti-HIV; mutant; mutein; immunogen.
XX
OS Human immunodeficiency virus 1; strain SF162 (clade B).
OS Synthetic.

Key Location/Qualifiers
FT Misc-difference 512
FT /note= "Encoded by CAGCTC"
XX
PN WO200200250-A2.
XX
PD 03-JAN-2002.
XX
PF 27-JUN-2001; 2001WO-US020483.
XX
PR 27-JUN-2000; 2000US-0214608P.
PR 26-JUN-2001; 2001US-00891609.
XX
PA (AARO-) AARON DIAMOND AIDS RES CENT.
XX

Stamatatos L, Barnett S, Shrivastava I;
WPI; 2002-130836/17.
DR N-PSDB; AAS15499.
XX

Immunizing an animal and eliciting an immune response against
PT heterologous HIV-1 in an animal, involves administering an immunogen
PT comprising modified HIV-1 envelope protein, or DNA or virus encoding the
PT protein.

XX Claim 5; Fig 17; 62pp; English.

XX

CC The present invention relates to methods for immunising an animal,
CC preferably human, against heterologous human immunodeficiency virus type
CC 1 (HIV-1), and eliciting a heterologous immune response to HIV-1 in that
CC animal. The method comprises administering an immunogen having at least
CC one modified HIV-1 envelope protein with a deletion in the second
CC hypervariable (V2) region or its fragment, or DNA or virus encoding the
CC modified HIV-1 envelope protein. The invention discloses the
CC polynucleotide and amino acid sequences for the full-length HIV-1
CC SF162deltaV2 (SF162DV2) gp140 envelope protein, and, for an N-terminal
CC mutant of SF162DV2. The immunogens of the invention are useful for
CC immunising an animal against heterologous HIV-1 strains by eliciting
CC neutralising antibodies or protective antibodies in the animal. The
CC method can be used to treat acquired immunodeficiency syndrome (AIDS).
CC The present sequence represents the N-terminal mutant (deletion of amino
CC acids 1-27) of modified HIV-1 SF162DV2 gp140 envelope protein
XX
SQ Sequence 619 AA;

Query Match 95.8%; Score 3327; DB 5; Length 619;
Best Local Similarity 100.0%; Pred. No. 5.8e-162;
Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 SAVEKLMVTYYGVVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 87
Db 1 SAVEKLMVTYYGVVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60

QY 88 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 147
Db 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

QY 148 DRGEIKNCSFKVGAGKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGS 207
Db 121 DRGEIKNCSFKVGAGKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGS 180

QY 208 GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVIRSENFDTNAKTIIVQLKESVEINC 267
Db 181 GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVIRSENFDTNAKTIIVQLKESVEINC 240

QY 268 TRPNNNTRKSITIGPRAFYATGDIIGDIRQAHCHNISGEKWNNTLKQIVTKLQAQFGNKT 327
Db 241 TRPNNNTRKSITIGPRAFYATGDIIGDIRQAHCHNISGEKWNNTLKQIVTKLQAQFGNKT 300

QY 328 IVFKOSSGGDPEIVMHSFNCGGEFFYCNSSTQLENSTWNNTIGPNTNGTITLPCRIKQII 387
Db 301 IVFKOSSGGDPEIVMHSFNCGGEFFYCNSSTQLENSTWNNTIGPNTNGTITLPCRIKQII 360

QY 388 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 447
Db 361 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 420

QY 448 YKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSILTLTVQARQ 507
Db 421 YKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSILTLTVQARQ 480

QY 508 LLSGIVQQONNLLRAIEAQHLLQLTVWGIKLQARVLAVERYLKDQQLLGIWCSGKLIC 567
Db 481 LLSGIVQQONNLLRAIEAQHLLQLTVWGIKLQARVLAVERYLKDQQLLGIWCSGKLIC 540

QY 568 TTAVPWNASWSNKSLDQIWNNTMTMEWEREIDNYTNLIYTLIEESQKQEKNEQELLELD 627
Db 541 TTAVPWNASWSNKSLDQIWNNTMTMEWEREIDNYTNLIYTLIEESQKQEKNEQELLELD 600

QY 628 KWASLWNWFDISKWLWYIK 646
Db 601 KWASLWNWFDISKWLWYIK 619

RESULT 4
ABR55684
ID ABR55684 standard; protein; 842 AA.
XX
AC ABR55684;
XX

Best Local Similarity 94.2%; Pred. No. 1.9e-159;			
Matches 619; Conservative 2; Mismatches 7; Indels 29; Gaps 3;			
Qy	19	LLGMLMIC-SAVEKLVTVVYGVVWKEATTLFCASDAKAYDEVHNVWATHACVPTD	77
Db	13	LLCGAVFVSPSAVEKLVTVVYGVVWKEATTLFCASDAKAYDEVHNVWATHACVPTD	72
Qy	78	PNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT	137
Db	73	PNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT	132
Qy	138	NTKSSNWKEMDRGEIKNCSEKVGAG-----KLINCNTS	170
Db	133	NTKSSNWKEMDRGEIKNCSEKVTTSIRNMQKEYALFYKLDVVPIDNDNTSYKLINCNTS	192
Qy	171	VITQACPVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQL	230
Db	193	VITQACPVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQL	252
Qy	231	LLNGSLAEBGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAYATG	290
Db	253	LLNGSLAEBGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAYATG	312
Qy	291	DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGE	350
Db	313	DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGE	372
Qy	351	FFYCNSTQLFNSTWNNTIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS	410
Db	373	FFYCNSTQLFNSTWNNTIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS	432
Qy	411	NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV	470
Db	433	NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV	492
Qy	471	QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL	530
Db	493	QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL	552
Qy	531	QLTVWGIK-LQARVLAVERYLKDQQLLGWCSGKLICTTAVPNWASWSNKSLDQIWNM	589
Db	553	QLTVWGIKQLQARVLAVERYLKDQQLLGWCSGKLICTTAVPNWASWSNKSLDQIWNM	612
Qy	590	TWMEWERIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK	646
Db	613	TWMEWERIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK	669

RESULT 6	
ABU66565	
ID	ABU66565 standard; protein; 842 AA.
XX	
AC	ABU66565;
XX	
DT	22-MAY-2003 (first entry)
XX	
DE	Human immunodeficiency virus (HIV) envelope (env) protein #1.
XX	
KW	Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW	gene therapy; packaging cell line; humoral immune response;
KW	cellular immune response; gene delivery vector; DNA immunisation;
KW	envelope protein; env.
XX	
OS	Human immunodeficiency virus.
XX	
PN	WO2003004657-A1.
XX	
PD	16-JAN-2003.
XX	
XX	05-JUL-2002; 2002WO-US021421.
XX	
PR	05-JUL-2001; 2001US-0303192P.
PR	31-AUG-2001; 2001US-0316860P.

PR	16-JAN-2002; 2002US-0349728P.		
PR	16-JAN-2002; 2002US-0349793P.		
PR	16-JAN-2002; 2002US-0349871P.		
XX			
PA	(CHIR) CHIRON CORP.		
XX			
PI	Zur Megede J, Barnett SW, Lian Y;		
XX			
DR	WPI; 2003-221602/21.		
XX			
PT	New synthetic polynucleotides encoding antigenic HIV type B and/or type C		
PT	polypeptides, useful as immunogenic compositions or vaccines for		
PT	generating humoral or cellular immune responses against HIV in a subject,		
PT	especially humans.		
XX			
PS	Disclosure; Fig 2A-C; 262pp; English.		
XX			
CC	The invention describes a synthetic polynucleotide encoding 2 or more		
CC	immunogenic HIV polypeptides, where at least 2 of the polypeptides are		
CC	derived from different HIV subtypes. The polynucleotide is useful for		
CC	immunisation. Generation of packaging cell lines, or production of HIV		
CC	polypeptides. The polynucleotide and its encoded proteins are useful as		
CC	immunogenic compositions or vaccines for generating humoral or cellular		
CC	immune responses against HIV in a subject, or for inducing neutralising		
CC	antibodies against HIV. The gene delivery vector comprising the		
CC	polynucleotide is also useful for DNA immunisation of, or for generating		
CC	an immune response (e.g. a humoral or cellular immune response) in, a		
CC	subject such as a mammal, particularly a human. This is the amino acid		
CC	sequence of a human immunodeficiency virus (HIV) envelope (env) protein		
XX			
SQ	Sequence 842 AA;		
Query Match 94.3%; Score 3274.5; DB 6; Length 842;			
Best Local Similarity 94.1%; Pred. No. 3.5e-159;			
Matches 618; Conservative 2; Mismatches 8; Indels 29; Gaps 3;			
Qy	19	LLGMLMIC-SAVEKLVTVVYGVVWKEATTLFCASDAKAYDEVHNVWATHACVPTD	77
Db	13	LLCGAVFVSPSAVEKLVTVVYGVVWKEATTLFCASDAKAYDEVHNVWATHACVPTD	72
Qy	78	PNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT	137
Db	73	PNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT	132
Qy	138	NTKSSNWKEMDRGEIKNCSEKVGAG-----KLINCNTS	170
Db	133	NTKSSNWKEMDRGEIKNCSEKVTTSIRNMQKEYALFYKLDVVPIDNDNTSYKLINCNTS	192
Qy	171	VITQACPVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQL	230
Db	193	VITQACPVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQL	252
Qy	231	LLNGSLAEBGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAYATG	290
Db	253	LLNGSLAEBGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAYATG	312
Qy	291	DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGE	350
Db	313	DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGE	372
Qy	351	FFYCNSTQLFNSTWNNTIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS	410
Db	373	FFYCNSTQLFNSTWNNTIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS	432
Qy	411	NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV	470
Db	433	NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV	492
Qy	471	QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL	530
Db	493	QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL	552
Qy	531	QLTVWGIK-LQARVLAVERYLKDQQLLGWCSGKLICTTAVPNWASWSNKSLDQIWNM	589

gp160 (env gene product) sequence. The invention provides compositions and methods for determining ancestral viral gene sequences and ancestral viral protein sequences for highly diverse viruses, such as HIV-1. The methods use samples of circulating viruses to determine an ancestral viral sequence by maximum likelihood phylogeny analysis. In the present case, the ancestral HIV-1 subtype B env sequence (see AAR26468) was determined using 38 subtype B sequences (obtained from 9 different countries) and 3 subtype D (outgroup) sequences. The distances between this ancestral viral sequence and circulating strains used to determine it were on average 12.3% (range: 8.0-21.0%) while the available specimens were 17.3% different from each other (range: 13.3-23.2%). Thus, the ancestor sequence was, on average, more closely related to any given circulating virus than to any other variant. The ancestral gp160 sequence included a wide variety of immunogenically active peptides when processed for antigen presentation; nearly all known subtype B CTL epitope consensus amino acids were represented. Thus, an immunogenic composition to this subtype B ancestor protein will elicit broad neutralising antibody against HIV-1 isolates of the same subtype, and will also elicit a broad cellular response mediated by antigen-specific T-cells. A claimed vaccine composition comprises a viral ancestor protein or its immunogenic fragment, especially one derived from the HIV-1 group M subtype B gp160 ancestral protein. (Updated on 11-SEP-2003 to standardise OS field)

QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNV 213
Db 181 PIDNDNTSVRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCRDKKFNGTGPCCTNV 240
QY 214 STVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNN 273
Db 241 STVQCTHGIRPVVSTQLLNGSLAEDEVVIRSANFSDNAKTIIVQLNESVEINCTRPNN 300
QY 274 TRKSITIGPGRAFYATGDIIGDIRQAHNCISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS 333
Db 301 TRRSIHIGPGRAFYATGEIIGDIRQAHNCNLSSTKWNTTLKQIVTKLREHF-NKTIVFNHS 359
QY 334 SGGDPEIVMHSFNCGGEFFYCNSSTQLFNS-----TWNNTIGPNNNTGTITLPCRICKQII 387
Db 360 SGGDPEIVMHSFNCGGEFFYCNTTPLFNSWTWNYTYTWNNTGSSNDTGRNITLQCRICKQII 419
QY 388 NRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTTEIFRPGGGDMRDNRSEL 447
Db 420 NMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGN--NSETEIFRPGGGDMRDNRSEL 477
QY 448 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 507
Db 478 YKYKVVKIEPLGVAPTAKRRVMQREKRAVGIGAVFLGFLGAAGSTMGAASVTLTVQARL 537
QY 508 LLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGIWCGSGKLI 566
Db 538 LLSGIVQQNNLLRAIEAEQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWCGSGKLI 597
QY 567 CTTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQKEQELLEL 626
Db 598 CTTAVPWNASWSNKSLDKIWDNMTWMEWEREIDNYTSLIYSLIEESQNQKEQELLEL 657
QY 627 DKWASLWNWFDISKWLWYIK 646
Db 658 DKWASLWNWFDITKWLWYIK 677

RESULT 10
ID ABR55495
XX ABR55495 standard; protein; 856 AA.
AC ABR55495;
XX
DT 11-AUG-2003 (first entry)
XX
DE Amino acid sequence of a HIV envelope protein.
XX
KW Polymorphic marker; host response; HIV; AIDS related virus; vaccine;
KW hepatitis related virus; HCV; HBV; drug resistance; envelope.
XX
OS Human immunodeficiency virus.
XX
PN WO2003035097-A1.
XX
PD 01-MAY-2003.
XX
PF 23-OCT-2002; 2002WO-AU001450.
XX
PR 23-OCT-2001; 2001AU-00008425.
XX
PA (EPIP-) EPIPOP PTY LTD.
XX
PI Mallal S;
XX
DR WPI; 2003-449231/42.
XX
PT Determining the influence of variation in host genes on the selection of
PT microorganisms with protein substitutions, comprises typing individuals
PT of a cohort infected with a microorganism for an intrinsic polymorphic
PT marker.
XX
PS Claim 22; Page 90-91; 157pp; English.

XX The specification describes a method of determining the influence of
CC variation in host genes on selection of microorganisms with protein
CC substitutions. The method comprises typing all individuals of a
CC population of patients infected with a microorganism for at least one
CC selected intrinsic polymorphic marker involved in the host response to
CC the presence of the microorganism. The method is useful for examining
CC selective pressures confronting a wide range of organisms that exhibit
CC pathogenic traits in a host, such as bacteria, fungi, mycobacterium,
CC viruses and virus-like particles; for examining microorganisms that have
CC adapted to evolve rapidly, including HIV and AIDS related viruses and the
CC hepatitis related viruses such as HCV and HBV. The method is useful for
CC designing a vaccine to prevent or delay the emergence of drug resistance
CC in patients treated with a particular drug specific for a microorganism
CC where the drug affects the replication of the microorganism at the
CC nucleotide or amino acid level. The present sequence represents a HIV
CC envelope protein, which is expected to provide optimal cytotoxic T
CC lymphocyte (CTL) induced therapeutic protection to the cohort examined in
CC that study
XX
SQ Sequence 856 AA;
Query Match 86.5%; Score 3006.5; DB 6; Length 856;
Best Local Similarity 84.2%; Pred. No. 1.6e-145;
Matches 579; Conservative 27; Mismatches 35; Indels 47; Gaps 11;
QY 1 MRVKGIRKNYQHL----WRGGTLLGLMLMICSAVEKLWTVVYGVVPVWKEATTLFCASD 56
Db 1 MRVKG--NNQHLWKGWKKGWMLGLMLMICSAATEKLWTVVYGVVPVWKEATTLFCASD 57
QY 57 AKAYDTEVHNWATHACVPTDPNPQEIIVLENTENFNMKNNVQMHEDIISLWDQSLK 116
Db 58 AKAYDTEVHNWATHACVPTDPNPQEVVLENTENFNMKNNVQMHEDIISLWDQSLK 117
QY 117 PCVKLTPLCVTLHCTNLKNATNT-KSSNWKEMDRGEIKNCSEFKVGAG----- 162
Db 118 PCVKLTPLCVTLNCTDLNNDTNTNTSGSNMKEGEIKNCSEFNITTSIRDKMQKEYALFY 177
QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSG 208
Db 178 KLDVVPIDNDNTSVRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGTG 237
QY 209 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCT 268
Db 238 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEDEVVIRSENFDTNAKTIIVQLNESVEINCT 297
QY 269 RPNNTRKSIIT--IGPGRAFYATGDIIGDIRQAHNCNISGEKWNNTLKQIVTKLQAQFG-N 325
Db 298 RPNNTRKSIISIHIGPGRAFYATGE-IGDIRQAHNCNISRAEWNNTLKQIVKKLREQFGKN 356
QY 326 KTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSSTQLFNSWTN---NTIGPNNNTNG--TITL 379
Db 357 KTIVFNQSSGGDPEIVMHSFNCGGEFFYCNTTQLFNSWTNNSWTWNTESNNTGNETITL 416
QY 380 PCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTTEIFRPGGGDM 439
Db 417 PCRIKQIINMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGNN--NNKTETFRPGGGDM 475
QY 440 RDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSL 499
Db 476 RDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGIGAMFLGFLGAAGSTMGAASI 535
QY 500 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGI 558
Db 536 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGI 595
QY 559 WGCSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQOEK 618
Db 596 WGCSGKLICTTAVPWNNTSWNSKSLNKIWDNMTWMEWEKEINNNTYGLIYNLIEESQNQOEK 655
QY 619 NEQELLELDKWASLWNWFDISKWLWYIK 646
Db 656 NEQELLELDKWASLWNWFDISKWLWYIK 683

Qy	269	RPNNNTRKSI	TIGPGRAF	YATGDI	IGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTI	328
Db	298	RPNNNTRKSI	TGPRVI	YATGQI	IGDIRKAHCNLSRAQWNNTLKQVVTKLREQFDNKTI	357
Qy	329	VFKQSSGDP	PEIVMHS	FCGGEFF	YCNSTQLFNSNTWNNTIGPNNTNG--TITLPCR	386
Db	358	VFTSSSGDP	PEIVLHS	FCGGEFF	YCNSTQLFNSNTWNNTIGPNNTNG--TITLPCR	417
Qy	387	INRWQEVGK	AMYPPI	RQIRCS	SNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSE	446
Db	418	VNMWQEVGK	AMYPPI	SGQIKCIS	NITGLLLTRDGGEDTNTTEIFRPGGDMRDNRSE	477
Qy	447	LYKYKVV	KIEPLG	VAPTAKRR	VVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQAR	506
Db	478	LYKYKVV	KIEPLG	VAPTAKRR	VVQREKRAVGIGALFLGFLGAAGSTMGAASMTLTVQAR	537
Qy	507	QLLSGIVQO	NNLLRA	IEAQOHL	LQLTVWGIK-LQARVLAVERYLKDQQLLGIWCSGKL	565
Db	538	QLLSGIVQO	NNLLRA	IEAQOHL	LQLTVWGIKQLOARILAVERYLKDQQLLGIWCSGKL	597
Qy	566	ICTTAVP	WNASWSN	KSLEQI	WNMTWMEWEREIDNYTNLIYTLIEESONQOEKNEQELLE	625
Db	598	ICTTAVP	WNASWSN	KSLEQI	WNMTWMEWDREINNYTSLIHSLEESONQOEKNEQELLE	657
Qy	626	LDKWASL	WNWFD	ISKWLWYIK	646	
Db	658	LDKWASL	WNWFEN	TNWLWYIK	678	
RESULT 12						
ID	AAW53112 standard; protein; 855 AA.					
AC	AAW53112;					
DT	17-OCT-2003 (revised)					
DT	25-JUN-1998 (first entry)					
XX	ENV protein contained in a complete ARV-2 nucleotide sequence.					
DE	ARV-2; enhanced promoter; gene expression; cytomegalovirus; HIV; AIDS;					
XX	ENV protein.					
KW	Human immunodeficiency virus 1.					
OS	US5688688-A.					
XX	18-NOV-1997.					
PN	10-AUG-1994; 94US-00288336.					
XX	31-OCT-1984; 84US-00667501.					
PR	30-JAN-1985; 85US-00696534.					
PR	06-SEP-1985; 85US-00773447.					
PR	24-DEC-1987; 87US-00138894.					
PR	17-AUG-1992; 92US-00931191.					
PR	28-JUN-1993; 93US-00083391.					
PR	17-AUG-1993; 93US-00107377.					
XX	(CHIR) CHIRON CORP.					
PA	Haigwood NL, Dina D, Rosenberg S, Chapman BS, Thayer RM;					
XX	Luciw PA;					
PI	WPI; 1998-007982/01.					
PI	N-PSDB; AAV04733.					
XX	Enhanced promoter for gene expression - comprising cytomegalovirus					
DR	immediate early promoter plus intron.					
DR	Example 1; Fig 4C-P; 99pp; English.					
XX	This sequence represents the ENV protein contained in a complete					
PS	CC					

Qy	269	RPNNNTRKSI	TIGPGRAF	YATGDI	IGDIRQA	HCNISGE	KWNNTLK	QIVTKL	QAQFGN	KTI	328	
Db	298	RPNNNTRKSI	TGPRVI	YATGQI	IGDIRKA	HCNLSRA	QWNNTLK	QVYTKL	REQFDN	KTI	357	
Qy	329	VFKOSSG	DP	PEIVMHS	FCGGEFF	YCNSTQ	LFNSTWN	NTIGPN	NTNG--	TITL	PCRIKQI	386
Db	358	VFTSSSG	DP	PEIVLHS	FCGGEFF	YCNSTQ	LFNSTWN	STEGSN	NTGNDT	TITL	PCRIKQI	417
Qy	387	INRWQEV	GKAMYAP	PIRQIR	CSNITGL	LLTRDGG	KEISNTTEI	FRPGG	DMRDN	WRSE	446	
Db	418	VNMWQEV	GKAMYAP	PISGQIK	CISNITGL	LLTRDGG	EDTTNTTEI	FRPGG	DMRDN	WRSE	477	
Qy	447	LYKYKV	VVKIEPL	GVAPT	KA	KRRVVQ	REKRAV	TGAMF	LGFLGA	AGSTMGAR	SLTLTVQAR	506
Db	478	LYKYKV	VVKIEPL	GVAPT	KA	KRRVVQ	REKRAV	GIGALF	LGFLGA	AGSTMGAAS	MTLTTVQAR	537
Qy	507	QLLSGIV	QQNNLLRA	IEAQO	QHLLQL	TVWG	IK-LQAR	VLAVER	YLKDQO	LLGIW	CSGKL	565
Db	538	QLLSGIV	QQNNLLRA	IEAQO	QHLLQL	TVWG	IKQLOAR	ILAVERY	LKDQO	LLGIW	CSGKL	597
Qy	566	ICTTAV	PNASWSN	KSLEQI	WNMTW	MEWERE	IDNYTN	LIYTLIE	SQOEK	NEQELLE	625	
Db	598	ICTTAV	PNASWSN	KSLEQI	WNMTW	MEWDRE	INNYTSL	HLIESQ	OEKNEQELLE	657		
Qy	626	LDKWAS	LWNW	FDISK	WLYIK	646						
Db	658	LDKWAS	LWNW	FNITN	WLYIK	678						
		RESULT 12										
		AAW53112										
ID		AAW53112										
XX												
AC												
XX												
DT												
DT												
XX												
DE												
XX												
KW												
KW												
XX												
OS												
XX												
PN												
XX												
PD												
XX												
PF												
XX												
PR												
PR												
XX												
PA												
XX												
PI												
XX												
DR												
DR												
XX												
PT												
PT												
XX												
PS												
XX												
CC												
CC												
CC												
CC												
CC												
CC												
CC												
XX												
SQ												
		Sequence 851 AA;										
		Query Match 86.3%; Score 2997; DB 1; Length 851;										
		Best Local Similarity 83.0%; Pred. No. 4.9e-145;										
		Matches 565; Conservative 40; Mismatches 38; Indels 38; Gaps 7;										
Qy	1	MRVKGIRKNYQHL	----	WRGGT	LLGLML	ICSAVE	KLWTV	VYVGV	PVWKEAT	TTLFCASD	56	
Db	1	MRVK---	EKYQHL	WRGWR	GTMLL	GLMLIC	SATEK	LWTV	VYVGV	PVWKEAT	TTLFCASD	57
Qy	57	AKAYDTEVHNVWATH	ACVPTD	PNPQEI	VLENVT	ENFNMW	KNNMVE	QMHEDI	ISLWDQSLK	116		
Db	58	AKAYDTEVHNVWATH	ACVPTD	PNPQEV	VLNVVT	ENFNMW	KNDMVE	QMHEDI	ISLWDQSLK	117		
Qy	117	PCVKLTPLCVTLHCT	NLKNAT	NKSNWKE	-MDRGEI	KNCSEF	KVUGA	----	GK-----	163		
Db	118	PCVKLTPLCVSLKCT	DLKNDT	NTNSSG	RNIMEK	GEIKNC	SFNMTS	IRGKVQ	EYAFFY	177		
Qy	164	-----	-----	LINCNTS	VITQAC	PKVSFE	PIPIHY	CAPAG	FAILKCN	DKKFN	GS	208
Db	178	KLDIIPIDNDT	TSYTLT	SCNTSVIT	QACPKV	SFEPIPIHY	CAPAG	FAILKCN	NKTFNGT	237		
Qy	209	PCTNVSTVOCTHGIR	PVWSTQ	LLNGSL	ABEGW	IRSENFT	DNAKTI	IVQLKE	SVEINCT	268		
Db	238	PCTNVSTVOCTHGIR	PVWSTQ	LLNGSL	ABEEW	IRSENFT	DNVKT	IVQLNA	SVQINCT	297		

Prodn. of recombinant HIV envelope proteins in insect cells - useful as vaccine against AIDS and for diagnosis and therapy.

Disclosure; Page ?; 4pp; English.

The sequence is the result of cloning a hybrid envelope gene from HIV variants BH10 and RF. A central portion of the RF gene was used, the rest being from te distantly related variant BH10. The resulting clone, pAHT6 produces a hybrid gp 160 envelope protein with novel immunological and antigenic characteristics. It may be used to as a vaccine and for diagnosis and therapy of AIDS. See also AAP80966. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)

Rusche J, Lynn D, Carson H, Putney S, Jellis CL;
WPI; 1988-176944/26.
N-PSDB; AAN80948.

HIV protein HT6.
HIV; HT6; gp160; envelope protein; RF; AIDS.

Human immunodeficiency virus; variant RF.
EP272858-A.

29-JUN-1988.
14-DEC-1987; 87EP-00310967.

15-DEC-1986; 86US-00941111.
31-AUG-1987; 87US-00091481.

(REPK) REPLIGEN CORP.

Enhanced promoter for gene expression - comprising cytomegalovirus immediate early promoter plus intron.
Example 1; Fig 4C-P; 99pp; English.

This sequence represents the ENV protein contained in a complete

ENV protein contained in a complete ARV-2 nucleotide sequence.

ARV-2; enhanced promoter; gene expression; cytomegalovirus; HIV; AIDS; ENV protein.

Human immunodeficiency virus 1.

US5688688-A.

18-NOV-1997.

10-AUG-1994; 94US-00288336.

31-OCT-1984; 84US-00667501.

30-JAN-1985; 85US-00696534.

06-SEP-1985; 85US-00773447.

24-DEC-1987; 87US-00138894.

17-AUG-1992; 92US-00931191.

28-JUN-1993; 93US-00083391.

17-AUG-1993; 93US-00107377.

(CHIR) CHIRON CORP.

Haigwood NL, Dina D, Rosenberg S, Chapman BS, Thayer RM; Luciw PA;

WPI; 1998-007982/01.

N-PSDB; AAV04733.

Enhanced promoter for gene expression - comprising cytomegalovirus immediate early promoter plus intron.

Example 1; Fig 4C-P; 99pp; English.

This sequence represents the ENV protein contained in a complete

CC nucleotide sequence of ARV-2 derived from partial sequences of several
CC ARV clones. The invention provides a method for construction of a vector
CC for expression of a polypeptide in a mammalian cell, comprising a
CC polypeptide coding sequence operably linked downstream of an enhanced
CC promoter. The enhanced promoter comprises the human cytomegalovirus
CC immediate early region (HCMV IE1) promoter and the first intron proximate
CC to the 3' end of the HCMV IE1 promoter. The polypeptide can be any of the
CC HIV recombinant polypeptides and especially HIV gp120. Expression of HIV
CC gp120 by COS 7 cells transfected with pCMV6a containing the gp120 coding
CC region, where pCMV6a is a vector containing the above enhanced promoter,
CC is increased by a factor of 50-100 compared with the use of a vector
CC containing the SV40 early promoter. (Updated on 17-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 855 AA;

AC AAY77298;
XX
XX 12-SEP-2003 (revised)
DT 22-MAY-2000 (first entry)
XX
XX HIV-1 (ATCC CRL 8597) env protein.
DE
XX
XX HIV-1; immunoassay; antigen; pol fragment; p31; recombinant; antibody;
KW detection; env protein.
XX
OS Human immunodeficiency virus 1; 'ATCC CRL 8597'.
XX
XX US6013432-A.
PN
XX
PD 11-JAN-2000.
XX
XX 17-MAY-1995; 95US-00443434.
PF
XX 31-OCT-1984; 84US-00667501.
PR 30-JAN-1985; 85US-00696534.
PR 06-SEP-1985; 85US-00773447.
PR 24-DEC-1987; 87US-00138894.
PR 17-AUG-1992; 92US-00931154.
PR 08-JUL-1993; 93US-00089407.
XX
XX (CHIR) CHIRON CORP.
PA
XX
XX Luciw PA, Dina D;
PI
XX
DR WPI; 2000-170256/15.
DR N-PSDB; AAZ90201.
XX
PT Immunoassay for antibodies against human immune deficiency virus, for
PT diagnosing infection, uses an immunogenic fragment of the pol protein as
PT antigen.
XX
PS Example 1; Fig 4K-O; 99pp; English.
PS
XX
CC The invention relates to the improvement of HIV-1 immunoassays by the use
CC of an HIV-1 antigen comprising an immunogenic fragment of recombinant or
CC synthetic HIV-1 pol, which is encoded by an approximately 9.7 kb sequence
CC between a BstXI restriction site at position 3006 and an NdeI site at
CC position 5131 of the genome (the proviral DNA sequence is given in
CC AAZ90201). The immunogenic pol fragment is not immunologically cross-
CC reactive with human T cell lymphotropic viruses I or II. The invention
CC also encompasses the use of p31 as an antigen. The recombinant antigens
CC may be produced in Escherichia coli, Saccharomyces cerevisiae or in
CC mammalian cells. Immunoassays using the recombinant HIV proteins may be
CC used to diagnose and stage HIV-1 infections. Sequences AAY7294-Y77299
CC represent proteins encoded by the genome of HIV-1 (ATCC CRL 8597).
CC (Updated on 12-SEP-2003 to standardise OS field)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:14:30 ; Search time 18.8949 Seconds
(without alignments)
1765.050 Million cell updates/sec

Title: US-09-891-609A-2
Perfect score: 3474
Sequence: 1 MRVKGIRKNYQHLLWRGGTLL.....DKWASLWNWFDISKWLWYIK 6466

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	3062.5	88.2	850	2	US-08-448-603A-28		Sequence 28, Appl
2	3062.5	88.2	850	3	US-09-134-075-28		Sequence 28, Appl
3	3062.5	88.2	850	4	US-09-492-739-28		Sequence 28, Appl
4	2997.5	86.3	855	3	US-07-956-483-14		Sequence 14, Appl
5	2978.5	85.7	855	1	US-08-022-835-6		Sequence 6, Appli
6	2978.5	85.7	855	1	US-08-388-809-6		Sequence 6, Appli
7	2978.5	85.7	855	2	US-08-647-714-6		Sequence 6, Appli
8	2976	85.7	855	3	US-07-956-483-15		Sequence 15, Appl
9	2974	85.6	857	2	US-08-448-603A-30		Sequence 30, Appl
10	2974	85.6	857	3	US-09-134-075-30		Sequence 30, Appl
11	2974	85.6	857	4	US-09-492-739-30		Sequence 30, Appl
12	2947	84.8	865	3	US-07-956-483-13		Sequence 13, Appl
13	2939.5	84.6	866	3	US-09-124-900-9		Sequence 9, Appli
14	2939.5	84.6	863	3	US-08-463-210-11		Sequence 11, Appl
15	2939.5	84.6	863	4	US-08-463-028-11		Sequence 11, Appl
16	2933	84.4	861	1	US-08-127-499A-14		Sequence 14, Appl
17	2933	84.4	861	1	US-08-482-847-14		Sequence 14, Appl
18	2933	84.4	861	3	US-07-956-483-10		Sequence 10, Appl
19	2933	84.4	861	3	US-08-472-240A-1		Sequence 1, Appli
20	2933	84.4	861	3	US-08-472-240A-7		Sequence 7, Appli
21	2933	84.4	861	4	US-08-817-441-103		Sequence 103, Appl
22	2931	84.4	867	3	US-08-472-240A-5		Sequence 5, Appli
23	2925	84.2	861	3	US-07-956-483-16		Sequence 16, Appl
24	2922.5	84.1	856	2	US-07-916-098A-2		Sequence 2, Appli
25	2919.5	84.0	854	4	US-09-309-572-23		Sequence 23, Appl
26	2919.5	84.0	854	4	US-09-718-096-23		Sequence 23, Appl
27	2913	83.9	880	2	US-08-788-815-7		Sequence 7, Appli

28	2813	83.9	880	3	US-09-157-963-7	Sequence 7, Appli
29	2907.5	83.7	856	4	US-09-337-387-11	Sequence 11, Appl
30	2904	83.6	839	3	US-08-472-240A-10	Sequence 10, Appl
31	2904	83.6	856	1	US-08-022-835-2	Sequence 2, Appli
32	2904	83.6	856	1	US-08-388-809-2	Sequence 2, Appli
33	2904	83.6	856	2	US-08-647-714-2	Sequence 2, Appli
34	2904	83.6	856	3	US-07-956-483-11	Sequence 11, Appl
35	2902	83.5	887	3	US-08-472-240A-6	Sequence 6, Appli
36	2900	83.5	857	1	US-08-022-835-4	Sequence 4, Appli
37	2900	83.5	857	1	US-08-388-809-4	Sequence 4, Appli
38	2900	83.5	857	2	US-08-647-714-4	Sequence 4, Appli
39	2886	83.1	726	4	US-09-337-387-3	Sequence 3, Appli
40	2882	83.0	759	4	US-09-337-387-12	Sequence 12, Appl
41	2879	82.9	856	1	US-08-375-100-1	Sequence 1, Appli
42	2879	82.9	887	3	US-08-472-240A-4	Sequence 4, Appli
43	2852	82.1	887	3	US-08-472-240A-2	Sequence 2, Appli
44	2843.5	81.9	665	2	US-08-448-603A-32	Sequence 32, Appl
45	2843.5	81.9	665	3	US-09-134-075-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
 US-08-448-603A-28
 ; Sequence 28, Application US/08448603A
 ; Patent No. 5864027
 ; GENERAL INFORMATION:
 ; APPLICANT: Berman, Phillip W.
 ; APPLICANT: Nakamura, Gerald R.
 ; TITLE OF INVENTION: HIV Envelope Polypeptides
 ; NUMBER OF SEQUENCES: 33
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
 ; STREET: 3 Embarcadero Center
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/448,603A
 ; FILING DATE: 07-JUN-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/072,833
 ; FILING DATE: 07-JUN-93
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haliday, Emily
 ; REGISTRATION NUMBER: 38903
 ; REFERENCE/DOCKET NUMBER: 14918-704
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-393-2000
 ; TELEFAX: 415-393-2286
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 28:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 850 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-448-603A-28

QY 1 MRVKGIRKNYQHILWRGGTLLGLMLMICS AVEKLMVTVVYGVVPWKEATTTLFCASDAKAY 60

Db 1 MIVKGIRKNCQHLWRWGTMLLGLMLMICSAAEKLWVTVYVGVPMKEATTTLFCASDAKAY 60
QY 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
QY 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEFKVGAG----- 162
Db 121 LTPLCVTLNCTDLKNATNTSSSWGKMERGEIKNCSEFNVTTSIRDKMKNEYALFYKLDVV 180
QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNV 213
Db 181 PIDNDNTSYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCRDXXKFNGTGPCTNV 240
QY 214 STVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNN 273
Db 241 STVQCTHGIRPVVSTQLLNGSLAEEEVVIRSANFSDNAKTIIVQLNESVEINCTRPNN 300
QY 274 TRKSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS 333
Db 301 TRRSIHIGPGRAFYATGEIIGDIRQAHCNLSSTKWNTTLKQIVTKLREHF-NKTIVFNHS 359
QY 334 SGGDPEIVMHSFNCGGEFFYCNTTPLFNSTWNNTYTYTNWNTGSGNDTGRNITLQCRIKQII 387
Db 360 SGGDPEIVMHSFNCGGEFFYCNTTPLFNSTWNNTYTYTNWNTGSGNDTGRNITLQCRIKQII 419
QY 388 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 447
Db 420 NMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGN--NSETEIFRPGGDMRDNRSEL 477
QY 448 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 507
Db 478 YKYKVVKIEPLGVAPTAKRRVMQREKRAVGIGAVFLGFLGAAGSTMGAASVTLTVQARL 537
QY 508 LLSGIVQOQNLLRAIEAQHLLQLTWGIK-LQARVLAVERYLKDQQLLGWCGSGKLI 566
Db 538 LLSGIVQOQNLLRAIEAEQHLLQLTWGIKQLQARVLAVERYLKDQQLLGWCGSGKLI 597
QY 567 CTTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQEKNEQELLEL 626
Db 598 CTTAVPWNASWSNKSLDKIWDNMTWMEWEREIDNYTSLIYSLIEESQNQEKNEQELLEL 657
QY 627 DKWASLWNWFDISKWLWYIK 646
Db 658 DKWASLWNWFDITKWLWYIK 677

RESULT 2
US-09-134-075-28
; Sequence 28, Application US/09134075
; Patent No. 6042836
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; APPLICANT: Nakamura, Gerald R.
; TITLE OF INVENTION: HIV Envelope Polypeptides
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
; STREET: 3 Embarcadero Center
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/448,603
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haliday, Emily
; REGISTRATION NUMBER: 38903
; REFERENCE/DOCKET NUMBER: 14918-704
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-393-2000
; TELEFAX: 415-393-2286
; TELEX:
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-134-075-28

Query Match 88.2%; Score 3062.5; DB 3; Length 850;
Best Local Similarity 85.3%; Pred. No. 9.1e-245;
Matches 580; Conservative 28; Mismatches 35; Indels 37; Gaps 5;

QY 1 MRVKGIRKNYQHLLWRGGTLLGLMLMICSABEKLWVTVYVGVPMKEATTTLFCASDAKAY 60
Db 1 MIVKGIRKNCQHLWRWGTMLLGLMLMICSAAEKLWVTVYVGVPMKEATTTLFCASDAKAY 60
QY 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
QY 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEFKVGAG----- 162
Db 121 LTPLCVTLNCTDLKNATNTSSSWGKMERGEIKNCSEFNVTTSIRDKMKNEYALFYKLDVV 180
QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNV 213
Db 181 PIDNDNTSYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCRDXXKFNGTGPCTNV 240
QY 214 STVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNN 273
Db 241 STVQCTHGIRPVVSTQLLNGSLAEEEVVIRSANFSDNAKTIIVQLNESVEINCTRPNN 300
QY 274 TRKSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS 333
Db 301 TRRSIHIGPGRAFYATGEIIGDIRQAHCNLSSTKWNTTLKQIVTKLREHF-NKTIVFNHS 359
QY 334 SGGDPEIVMHSFNCGGEFFYCNTTPLFNSTWNNTYTYTNWNTGSGNDTGRNITLQCRIKQII 387
Db 360 SGGDPEIVMHSFNCGGEFFYCNTTPLFNSTWNNTYTYTNWNTGSGNDTGRNITLQCRIKQII 419
QY 388 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 447
Db 420 NMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGN--NSETEIFRPGGDMRDNRSEL 477
QY 448 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 507
Db 478 YKYKVVKIEPLGVAPTAKRRVMQREKRAVGIGAVFLGFLGAAGSTMGAASVTLTVQARL 537
QY 508 LLSGIVQOQNLLRAIEAQHLLQLTWGIK-LQARVLAVERYLKDQQLLGWCGSGKLI 566
Db 538 LLSGIVQOQNLLRAIEAEQHLLQLTWGIKQLQARVLAVERYLKDQQLLGWCGSGKLI 597
QY 567 CTTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQEKNEQELLEL 626
Db 598 CTTAVPWNASWSNKSLDKIWDNMTWMEWEREIDNYTSLIYSLIEESQNQEKNEQELLEL 657
QY 627 DKWASLWNWFDISKWLWYIK 646
Db 658 DKWASLWNWFDITKWLWYIK 677

RESULT 3
US-09-492-739-28

QY	1	MRVKGIRKNYQHLWRGGTLLGLMLMI CSAVEKLVVTVYVYGVPMWKEATTTLFCASDAKAY	60
Db	1	MRVKGSGRNYQHLWRWGTMLLGIILMI CSAAEQLVVTVYVYGVPMWKEATTTLFCASDAKAY	60
QY	61	DTEVHNWATHACVPTDPNPOEIVLENVTENFNWKNMVEQMHEDIISLWDQSLKPCVK	120
Db	61	DTEVHNIWATHACVPTDPNPOEVVLGNVTENFNWKNMVEQMHEDIISLWDQSLKPCVK	120
QY	121	LTPLCVTLHCTNLK-----NATNTKSSNWKEMDRGEIKNCSEFKVGAG-----	162
Db	121	LTPLCVTLNCTNLRNDTSTNATNTSSNRGKMEGGEMTNCSEFNITTSIRSKVQKEYALFY	180
QY	163	-----KLINCNTSVITQACPKVSPEPIPIHYCAPAGFAILKCNCKKFNKGGPC	210
Db	181	KLDVVPIDNTSYTLINCNTSVITQACPKVSPEPIPIHYCA-RWFAILNCNNKKFNKGTGPC	239
QY	211	TNVSTVQCTHGIRPVVSTQLLNGSLAEEGVWIRSENFDTNAKTIIVQLKESVEINCTRP	270
Db	240	TNVSTVQCTHGIRPVVSTHLLNGSLAEEEVWLRSENFDTNAKTIIVQLKEAVEINCTRP	299
QY	271	NNNTRKSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIIV	330
Db	300	NNNTTRSIIHGPGRAFYATGDIIGDIRQAHCNISRAKWNNTLKQIVTKLRDQFENKTIIF	359
QY	331	KQSSGGDPEIIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNTNG--TITLPCRKIQIIN	388
Db	360	NRSSGGDPEIIVMHSFNCGGEFFYCNSTQLFSSTWNGTEGSNNTGNDTITLPCRKEIIN	419
QY	389	RWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNT----TEIFRPGGGDMRDNR	444
Db	420	MWQEVGKAMYAPPPIKGQVKCSSNITGLLLTRDGGNSKNGSKNENTEIFRPGGGDMRDNR	479
QY	445	SELYKYKWKIEPLGVAPTAKRRVVQREKRAV-TLGAMFLGFLGAAGSTMGARSLTLTV	503
Db	480	SELYKYKWKIEPLGVAPTAKRRVVQREKRAVGTIGAMFLGFLGAAGSTMGATSMTLTV	539
QY	504	QARQLLSGIVQQONLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDOQLLGIWGCS	562
Db	540	QARLLSGIVQQONLLRAIEAQHLLQLTVWGIKQLQARVLAVERYLDRDQQLGIWGCS	599
QY	563	GKLICTTAVPWNASWSNKSLDQIWNNTWMWEWEREIDNYTNLIYTLIEESNQOEKNEQE	622
Db	600	GKLICTTVPWNTSWSNKSLDKIWGNNTWMWEWEREIDNYTSLIYTLIEESNQOEKNEQE	659
QY	623	LLELDKWASLWNWFDISKWLWYIK	646
Db	660	LLELDKWASLWNWFNITNLWLWYIK	683

RESULT 5

US-08-022-835-6
; Sequence 6, Application US/08022835
; Patent No. 5420030

GENERAL INFORMATION:

APPLICANT: Reitz Jr., Marvin S.
APPLICANT: Franchini, Genoveffa
APPLICANT: Markham, Phillip D.
APPLICANT: Gallo, Robert C.
APPLICANT: Lori, Franco C.
APPLICANT: Popovic, Mikulas
APPLICANT: Garnter, Suzanne

; AFRICANI; GAINCEL; SUZANNE
 ;
 ; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES
 ; TITLE OF INVENTION: THEREOF
 ; NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: CUSHMAN, DARBY & CUSHMAN
STREET: Eleventh Floor, 1615 L. Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-5601
COMPUTER READABLE FORM:


```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 855 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-647-714-6

Query Match      85.7%; Score 2978.5; DB 2; Length 855;
Best Local Similarity 82.3%; Pred. No. 8.2e-238;
Matches 561; Conservative 36; Mismatches 48; Indels 37; Gaps 5;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICSABEKLWVTYYGVPVWKEATTTLFCASDAKAY 60
Db 1 MRVTEIRKSYQHWWRGIMLLGILMICNABEKLWVTYYGVPVWKEATTTLFCASDRKAY 60

Qy 61 DTEVHNWATHACVPTDPNPQEIIVLENVTFENFMWKNNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWATHACVPTDPNPQEVLELKNVTENFMWKNNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDR-----GEIKNCSFKVGAG----- 162
Db 121 LTPLCVTLNCTDLRNATNGNDTNTSSRGVMVGGMKNCSFNITTNIRKVKQKEYALFY 180

Qy 163 -----KLINCNTSVITQACPVKSFPEPIPIHYCAPAGFAILKCNDDKKFNGS 207
Db 181 KLDIAPIDNNSNNRYRLISCNTSVITQACPVKSFPEPIPIHYCAPAGFAILKCKDKKFNGK 240

Qy 208 GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVRSENFTDNAKTIIVQLKESVEINC 267
Db 241 GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEEVWIRSANFADNAKVIIVQLNESVEINC 300

Qy 268 TRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAFGNKT 327
Db 301 TRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCNLSRAKWNNTLNKIVIKLREQFGNKT 360

Qy 328 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNT--NGTITLPCRICKQ 385
Db 361 IVFKHSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWNVTEESNNVTENNTITLPCRICKQ 420

Qy 386 IINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNWS 445
Db 421 IINMWQEVGRAMYAPPPIRGQIRCSSNITGLLLTRDGGPE-DNKTEVFRPGGDMRDNWS 479

Qy 446 ELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAGSTMGARSLTLTVQA 505
Db 480 ELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGIGAVFLGFLGAGSTMGAAMTLTVQA 539

Qy 506 RQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGWGCSGK 564
Db 540 RLLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQARVLAVERYLRDQQLLGWGCSGK 599

Qy 565 LICTTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQKEKNEQELL 624
Db 600 LICTTAVPWNASWSNKS LNKIWDNMTWIEWREINNYTSIIYSLIEESQNQKEKNEQELL 659

Qy 625 ELDKWASLWNWEDISKWLWYIK 646
Db 660 ELDKWASLWNWEDITKWLWYIK 681

RESULT 8
US-07-956-483-15
; Sequence 15, Application US/07956483
; Patent No. 6261799
; GENERAL INFORMATION:
; APPLICANT: KIENY, Marie-Paule
; TITLE OF INVENTION: NOVEL HYBRID, SOLUBLE AND UNCLEAVED
; TITLE OF INVENTION: gp160 VARIANT

```

```

; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/956,483
; FILING DATE: 31-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/19742
; FILING DATE: 12-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 05392
; FILING DATE: 02-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 017753-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 855 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-956-483-15

Query Match      85.7%; Score 2976; DB 3; Length 855;
Best Local Similarity 82.3%; Pred. No. 1.3e-237;
Matches 562; Conservative 34; Mismatches 49; Indels 38; Gaps 6;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICSABEKLWVTYYGVPVWKEATTTLFCASDAKAY 60
Db 1 MKVKSTFRNYYQHLWRWGTTLLGLMLMICSATEKLWVTYYGVPVWKEATTTLFCASDARAY 60

Qy 61 DTEVHNWATHACVPTDPNPQEIIVLENVTFENFMWKNNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWATHACVPTDPNPQEVVGLNVTFENFMWKNNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db 121 LTPLCVTLNCTDLGKATNTNSSNWKEIKGEIKNCSFNITTSIRDKIQKENALFRNLDVV 180

Qy 163 -----KLINCNTSVITQACPVKSFPEPIPIHYCAPAGFAILKCNDDKKFNGSG 208
Db 181 PIDNASTTNTYRLIHCNRSVITQACPVKSFPEPIPIHYCTPAGFAILKCNKKTENGKG 240

Qy 209 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVRSENFTDNAKTIIVQLKESVEINCT 268
Db 241 TCTNVSTVQCTHGIRPIVSTQLLNGSLAEEEVWIRSDNFTNNAKTIIVQLNESVAINCT 300

Qy 269 RPNNNTKRSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAFQF-NKT 327
Db 301 RPNNNTKRSIYIGPGRAFHTTGTGRIIGDIRKAHCNISRAQWNNTLEQIVKKLREQFGNKT 360

Qy 328 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTW--NNTIGPNNTNGTITLPCRICKQ 385
Db 361 IVFNQSSGGDPEIVMHSFNCRCGEFFYCNSTQLFNNTWRLNHTTEGFKG-NDTITLPCRICKQ 419

Qy 386 IINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNWS 445

```



```

;
;   REGISTRATION NUMBER: 38903
;   REFERENCE/DOCKET NUMBER: 14918-704
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-393-2000
;   TELEFAX: 415-393-2286
;   TELEX:
;
;   INFORMATION FOR SEQ ID NO: 30:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 857 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;
US-09-134-075-30

Query Match      85.6%; Score 2974; DB 3; Length 857;
Best Local Similarity 81.2%; Pred. No. 1.9e-237;
Matches 556; Conservative 43; Mismatches 46; Indels 40; Gaps 4;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICS AVEK L WTVVYGV P V W K E A T T T L F C A S D A K A Y 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MRVKGIRRNYYQHLWRWGTMLLGILMICS AAGK L WTVVYGV P V W K E T T T T L F C A S D A K A Y 60

Qy 61 DTEVHNWATHACVPTDPNPQEI VLE NVTFENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 DTEIHNWATHACVPTDPNPQEV VLE NVTFENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 LTPLCVTLNCTDAGNTTNTSSSREKLEKGEIKNC SFNITTSVRDKMQKETALFNKLDIV 180

Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFG 206
   :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 PIDDDRNSTRNSTNYRLISCN TSVITQACPKVSFEPIPIHFCTPAGFALLKCNKKTFG 240

Qy 207 SGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE G VVIRSENFNTD NAKTIIVQLKESVEIN 266
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 SGPCKNVSTVQCTHGIRPVVSTQLLNGSLAE G EVVIRSENFNTN AKTIIVQLTEPVKIN 300

Qy 267 CTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNK 326
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 CTRPNNNTRKSIPIGPGRAFYATGDIIGNIRQAHCNLSR TDWNTKVSNGTSTEENSTITLPCR 420

Qy 327 TIVFKQSSGGDPEIVMHSFNCRGEFFYCNSTQLFNSTWNTTIGPNNT----NGTITLPCR 382
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 361 TIIFNHSSGGDPEIVMHSFNCRGEFFYCN TQLFDSTWDNTKVSNGTSTEENSTITLPCR 420

Qy 383 IKQIINRWQEVGKAMYAPPIRQIIRCSSNITGLLLTRDGGKEISNTTTEIFRPGGGDMRDN 442
   ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 421 IKQIVNMWQEVGKAMYAPPIRQIIRCSSNITGLLLTRDGGSN-NSMNETFRPGGGDMRDN 479

Qy 443 WRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLT 502
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 480 WRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGIGAVFLGFLGAAGSTMGAASITLT 539

Qy 503 VQARQLLSGIVQQNNLLRAIEAQHLLQLT VWGIK-LQARVLAVERYLKDQQLLGIWGC 561
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 540 VQARLLLSGIVQQNNLLRAIEAQHLLQLI VWGIKQLQARVLAVERYLRDQQLLGIWGC 599

Qy 562 SGKLICTTAVPNWASWSNKS L DQIWNNTWMEWEREIDNYTNLIYTLIEESQNQQEKNEQ 621
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 600 SGKLICTTSPWNASWSNKS LDKIWDNMTWMEWEREIENTYTSLIYTLIEESQNQQEKNEQ 659

Qy 622 ELLELDKWASLWNWFDISKWLWYIK 646
   :||||:|||||:|||||:|||||
Db 660 DLELDQWASLWNWFSITKWLWYIK 684

RESULT 11
US-09-492-739-30
; Sequence 30, Application US/09492739
; Patent No. 6331404
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
```

```

;
;   Nakamura, Gerald R.
;   TITLE OF INVENTION: HIV Envelope Polypeptides
;   NUMBER OF SEQUENCES: 33
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
;   STREET: 3 Embarcadero Center
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94111
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/492,739
;   FILING DATE: 27-Jan-2000
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/134,075
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Haliday, Emily
;   REGISTRATION NUMBER: 38903
;   REFERENCE/DOCKET NUMBER: 14918-704
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-393-2000
;   TELEFAX: 415-393-2286
;   TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 30:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 857 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 30:
;
US-09-492-739-30

Query Match      85.6%; Score 2974; DB 4; Length 857;
Best Local Similarity 81.2%; Pred. No. 1.9e-237;
Matches 556; Conservative 43; Mismatches 46; Indels 40; Gaps 4;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICS AVEK L WTVVYGV P V W K E A T T T L F C A S D A K A Y 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MRVKGIRRNYYQHLWRWGTMLLGILMICS AAGK L WTVVYGV P V W K E T T T T L F C A S D A K A Y 60

Qy 61 DTEVHNWATHACVPTDPNPQEI VLE NVTFENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 DTEIHNWATHACVPTDPNPQEV VLE NVTFENFNMWKNMVEQM HEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 LTPLCVTLNCTDAGNTTNTSSSREKLEKGEIKNC SFNITTSVRDKMQKETALFNKLDIV 180

Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFG 206
   :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 PIDDDRNSTRNSTNYRLISCN TSVITQACPKVSFEPIPIHFCTPAGFALLKCNKKTFG 240

Qy 207 SGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE G VVIRSENFNTD NAKTIIVQLKESVEIN 266
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 SGPCKNVSTVQCTHGIRPVVSTQLLNGSLAE G EVVIRSENFNTN AKTIIVQLTEPVKIN 300

Qy 267 CTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNK 326
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 CTRPNNNTRKSIPIGPGRAFYATGDIIGNIRQAHCNLSR TDWNTKVSNGTSTEENSTITLPCR 420

Qy 383 IKQIINRWQEVGKAMYAPPIRQIIRCSSNITGLLLTRDGGKEISNTTTEIFRPGGGDMRDN 442
   ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 421 IKQIVNMWQEVGKAMYAPPIRQIIRCSSNITGLLLTRDGGSN-NSMNETFRPGGGDMRDN 479

Qy 443 WRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLT 502
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 480 WRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGIGAVFLGFLGAAGSTMGAASITLT 539

Qy 503 VQARQLLSGIVQQNNLLRAIEAQHLLQLT VWGIK-LQARVLAVERYLKDQQLLGIWGC 561
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 540 VQARLLLSGIVQQNNLLRAIEAQHLLQLI VWGIKQLQARVLAVERYLRDQQLLGIWGC 599

Qy 562 SGKLICTTAVPNWASWSNKS L DQIWNNTWMEWEREIDNYTNLIYTLIEESQNQQEKNEQ 621
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 600 SGKLICTTSPWNASWSNKS LDKIWDNMTWMEWEREIENTYTSLIYTLIEESQNQQEKNEQ 659

Qy 622 ELLELDKWASLWNWFDISKWLWYIK 646
   :||||:|||||:|||||:|||||
Db 660 DLELDQWASLWNWFSITKWLWYIK 684
```


Db 421 IKQIVNMWQEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGSN-NSMNETFRPGGDMRDN 479
QY 443 WRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAAGSTMGARSLTLT 502
Db 480 WRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGIGAVFLGLGAAGSTMGAASITLT 539
QY 503 VQARQLLSGIVQQNNLLRAIEAQHLLQLTWGIK-LQARVLAVERYLKDQQLLGWGC 561
Db 540 VQARLLLSGIVQQNNLLRAIEAQHLLQLIWGIKQLQARVLAVERYLRDQQLLGWGC 599
QY 562 SGKLICTTAVPWNASWSNKS LDOIWNMTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQ 621
Db 600 SGKLICTTSVPWNASWSNKS LDKIWDNMTWMEWEREIEYNTSLIYTLIEESQNQOEKNEQ 659
QY 622 ELLELDKWASLWNWFDISKWLWYIK 646
Db 660 DLELDQWASLWNWFSITKWLWYIK 684

RESULT 12
US-07-956-483-13
; Sequence 13, Application US/07956483
; Patent No. 6261799
; GENERAL INFORMATION:
; APPLICANT: KIENY, Marie-Paule
; TITLE OF INVENTION: NOVEL HYBRID, SOLUBLE AND UNCLEAVED
; TITLE OF INVENTION: 9p160 VARIANT
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/956,483
; FILING DATE: 31-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/19742
; FILING DATE: 12-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 05392
; FILING DATE: 02-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 017753-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 865 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-956-483-13

Query Match 84.8%; Score 2947; DB 3; Length 865;
Best Local Similarity 80.1%; Pred. No. 3.4e-235;
Matches 554; Conservative 40; Mismatches 52; Indels 46; Gaps 5;
QY 1 MRVKGIRKQYQHLWRGGTLLGLMLICSAVEKLWTVVYGVVWKEATTLFCASDAKAY 60
Db 1 MRVMEMRKNCQHLWKWGTMLLGLMLICSAEADLWTVVYGVVWKEATTLFCASEAKAY 60

QY 61 DTEVHNWATHACVPTDPNPQEIIVLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 KTEVHNWAKHACVPTDPNPQEVLLLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
QY 121 LTPLCVTLHCT--NLKNATNTKSSNWKEMDRGEIKNCSFKV----- 159
Db 121 LTPLCVTLNCTDANLNGTNTVTSSSGGTMMENGEIKNCSFQVTTSSRRDKTKQKYALFYKLD 180
QY 160 -----GAGKLINCNTSVITQACPKVSFEPPIHYCAPAGFAILK 198
Db 181 VVPIEKGNISPKNNTSNNTSYGNYTLIHCNSSVITQACPKVSFEPPIHYCTPAGFAILK 240
QY 199 CNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNNAKTIIVQ 258
Db 241 CNDKKFNGTGPKNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNNAKTIIVQ 300
QY 259 LKESVEINCTRPNNNTRKSTIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTK 318
Db 301 LNASVQINCTRPNNNTRKSTIGPGRVIYATGQIIGDIRKAHCNLSRAQWNNNTLKQVVTK 360
QY 319 LQAFQGNKTIIVFKQSGGDPPIVMHSFNCGGEFFYCNSTQLPNSTWNTTIGPNNNTNG--T 376
Db 361 LREQFDNKTIIVFTSSSGGDPPIVLHSFNCGGEFFYCNSTQLPNSTWNTSTEGSNNTGNDT 420
QY 377 ITLPCRKIQIINRWQEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGG 436
Db 421 ITLPCRKIQIVNMWQEVGKAMYAPPISGQIKCISNITGLLLTRDGGEDTTNTTEIFRLGG 480
QY 437 GMDRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAV-TLGAMFLGLGAAGSTMG 495
Db 481 GMDRDNWRSELYKYKVVRIEPLGVAPTAKRRVVQREKRAVGTIGAMFLGLGAAGSTMG 540
QY 496 ARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTWGIK-LQARVLAVERYLKDQ 554
Db 541 AGSITLTVQARHLLSGIVQQNNLLRAIEAQHLLQLTWGIKQLQARVLAVERYLRDQ 600
QY 555 LLGIWCGSGKLICTTAVPWNASWSNKS LDOIWNMTWMEWEREIDNYTNLIYTLIEESQN 614
Db 601 LLGIWCGSGKLICTTTPWNASWSNKS LNMIMNMTWQWEREIDNYTGIIYNLLEESQN 660
QY 615 QOEKNEQELLELDKWASLWNWFDISKWLWYIK 646
Db 661 QOEKNEQELLELDKWANLWNWFDITQWLWYIR 692

RESULT 13
US-09-124-900-9
; Sequence 9, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124,900
; PRIOR FILING DATE: 1998-07-30
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 9
; LENGTH: 856
; TYPE: PRT

```

; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (79)..(184)
; NAME/KEY: PEPTIDE
; LOCATION: (326)..(400)
US-09-124-900-9

Query Match      84.6%; Score 2939.5; DB 3; Length 856;
Best Local Similarity 82.6%; Pred. No. 1.4e-234;
Matches 568; Conservative 34; Mismatches 39; Indels 47; Gaps 12;

Qy 1 MRVKGIRKNYQHL-----WRGGTLLGLMLMICS AVEKLVTVVYGVVPVWKEATTTLFCASD 56
    ||||| : ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MRVK---EKYQHLWRGWRWGTMLLGLMLMICS ATEKLVTVVYGVVPVWKEATTTLFCASD 57

Qy 57 AKAYDTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQM HEDIISLWDQSLK 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 AKAYDTEVHNWATHACVPTDPNPQEVVL VNVTFENFMWKNDMVEQM HEDIISLWDQSLK 117

Qy 117 PCVKLTPLCVTLHCTNLKNA TNTKSSNWKE-MDRGEIKNCSFKVGA---GK----- 163
    ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 PCVKLTPLCVSLKCTDLKNDTNTNSSGRMIMEKEIKNCSFNISTSI RGVKEYAFFY 177

Qy 164 -----LINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCN DKKFNCGS 208
    ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 KLDIIPIDNDTTSYTLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCN NKTFTNGTG 237

Qy 209 PCTNVSTVQCTHGIRPVVSTQLLNGSLAE EGVVIRSENFNDNAKTIIVQLKESVEINCT 268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 PCTNVSTVQCTHGIRPVVSTQLLNGSLAE EEEVIRSANFTDNAKTIIVQLNQSV EINCT 297

Qy 269 RPNNTRKSITI--GPGRAFYATGDIIGDIRQAH CNISGEKWNNTLKQIVTKLQAQFG-N 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 RPNNTRKSIRIQRGPGRAFVTIGK-IGNMRQAHCNISRAKWNNTLKQIDSKLREQFGNN 356

Qy 326 KTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNN----TIGPNNTNG--TITL 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 KTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTKGSNNTGSDTITL 416

Qy 380 PCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDM 439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 PCRIKQIINRWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNS--NNESEIFRPGGGDM 475

Qy 440 RDNWRSELYKYVVKVIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAAGSTMGARSL 499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 RDNWRSELYKYVVKVIEPLGVAPTAKRRVVQREKRAVGIGALFLGLGAAGSTMGAASM 535

Qy 500 TLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWG I K-LQARVLAVERYLKDQQLLGI 558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 536 TLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWG I KQLQARILAVERYLKDQQLLGI 595

Qy 559 WCGSGKLICTTAVPWNASWSNKS LDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQOEK 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 596 WCGSGKLICTTAVPWNASWSNKSLEQIWNMTWMEWDREINNYTSLIHS LIEESQNQOEK 655

Qy 619 NEQELLELDKWASLWNWFDISKWLWYIK 646
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 656 NEQELLELDKWASLWNWFNITNLWLWYIK 683

RESULT 14
US-08-463-210-11
; Sequence 11, Application US/08463210
; Patent No. 6001977
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
```

```

; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,210
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 863 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..863
; OTHER INFORMATION: /note= "env protein of HTLV-III"
US-08-463-210-11

Query Match      84.6%; Score 2939.5; DB 3; Length 863;
Best Local Similarity 82.6%; Pred. No. 1.4e-234;
Matches 568; Conservative 34; Mismatches 39; Indels 47; Gaps 12;

Qy 1 MRVKGIRKNYQHL-----WRGGTLLGLMLMICS AVEKLVTVVYGVVPVWKEATTTLFCASD 56
    ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8 MRVK---EKYQHLWRGWRWGTMLLGLMLMICS ATEKLVTVVYGVVPVWKEATTTLFCASD 64

Qy 57 AKAYDTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQM HEDIISLWDQSLK 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 AKAYDTEVHNWATHACVPTDPNPQEVVL VNVTFENFMWKNDMVEQM HEDIISLWDQSLK 124

Qy 117 PCVKLTPLCVTLHCTNLKNA TNTKSSNWKE-MDRGEIKNCSFKVGA---GK----- 163
    ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 PCVKLTPLCVSLKCTDLKNDTNTNSSGRMIMEKEIKNCSFNISTSI RGVKEYAFFY 184

Qy 164 -----LINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCN DKKFNCGS 208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185 KLDIIPIDNDTTSYTLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCN NKTFTNGTG 244

Qy 209 PCTNVSTVQCTHGIRPVVSTQLLNGSLAE EGVVIRSENFNDNAKTIIVQLKESVEINCT 268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 245 PCTNVSTVQCTHGIRPVVSTQLLNGSLAE EEEVIRSANFTDNAKTIIVQLNQSV EINCT 304

Qy 269 RPNNTRKSITI--GPGRAFYATGDIIGDIRQAH CNISGEKWNNTLKQIVTKLQAQFG-N 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 305 RPNNTRKSIRIQRGPGRAFVTIGK-IGNMRQAHCNISRAKWNNTLKQIDSKLREQFGNN 363

Qy 326 KTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNN----TIGPNNTNG--TITL 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 364 KTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSQLFNSTWFSNWTSGSNNTGSDTITL 423
Qy 380 PCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDM 439
Db 424 PCRIKQIINMWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNS--NNESEIFRPGGGDM 482
Qy 440 RDNWRSELYKYVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSL 499
Db 483 RDNWRSELYKYVVKIEPLGVAPTAKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASM 542
Qy 500 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGI 558
Db 543 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQARILAVERYLKDQQLLGI 602
Qy 559 WCGSGKLICTTAVPWNASWSNKSLEQIWNNTWMEWEREIDNYTNLIYTLIEESQNQOEK 618
Db 603 WCGSGKLICTTAVPWNASWSNKSLEQIWNNTWMEWDREINNYTSLIHSLEESQNQOEK 662
Qy 619 NEQELLELDKWASLWNWFDISKWLWYIK 646
Db 663 NEQELLELDKWASLWNWFNITNLWLWYIK 690

RESULT 15
US-08-463-028-11
; Sequence 11, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,028
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 863 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:

; NAME/KEY: Protein
; LOCATION: 1..863
; OTHER INFORMATION: /note= "env protein of HTLV-III"
US-08-463-028-11
Query Match 84.6%; Score 2939.5; DB 4; Length 863;
Best Local Similarity 82.6%; Pred. No. 1.4e-234;
Matches 568; Conservative 34; Mismatches 39; Indels 47; Gaps 12;
Qy 1 MRVKGIRKKNYQHL-----WRGGTLLGLMLMICSAVEKLWVTYYGVVPWKEATTTLFCASD 56
Db 8 MRVK---EKYQHLWRGWRGTMLLGLMICSAATEKLWVTYYGVVPWKEATTTLFCASD 64
Qy 57 AKAYDEVHNVWATHACVPTDPNPQEIIVLENTENFNMKNMVEQMHEDIISLWDQSLK 116
Db 65 AKAYDEVHNVWATHACVPTDPNPQEVVLNVNTENFNMKNMVEQMHEDIISLWDQSLK 124
Qy 117 PCVKLTPLCVTLHCTNLKNATNTKSSNWKE-MDRGEIKNCSPKVGA---GK----- 163
Db 125 PCVKLTPLCVSLKCTDLKNDTNTNTSSSGRMIMEKEIKNCSEFNISTSIIRGKVQKEYAFFY 184
Qy 164 -----LINCNTSVITQACPKVSEPIPIHYCAPAGFAILKCNKKTFTNGTG 244
Db 185 KLDIIPIDNDTTSYTLTSCNTSVITQACPKVSEPIPIHYCAPAGFAILKCNKKTFTNGTG 244
Qy 209 PCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEGWIRSENFTDNTAKTIIVQLKESVEINCT 268
Db 245 PCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEVWIRSANFTDNTAKTIIVQLNQSVINCT 304
Qy 269 RPNNTTRKSITI--GPGRAFATGDIIGDIRQAHNCNISGEKNNNTLKQIVTKLQAOFG-N 325
Db 305 RPNNTTRKSIRIQRGPGRAFTVIGK-IGNMRQAHNCNISRAKWNNTLKQIDSKLREQFGNN 363
Qy 326 KTIIFKQSSGGDPEIIVHSHFNCGGGEFFYCNSQLFNSTWNN-----TIGPNTNG--TITL 379
Db 364 KTIIFKQSSGGDPEIIVHSHFNCGGGEFFYCNSQLFNSTWFSNWTSGSNNTGSDTITL 423
Qy 380 PCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDM 439
Db 424 PCRIKQIINMWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNS--NNESEIFRPGGGDM 482
Qy 440 RDNWRSELYKYVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSL 499
Db 483 RDNWRSELYKYVVKIEPLGVAPTAKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASM 542
Qy 500 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGI 558
Db 543 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQARILAVERYLKDQQLLGI 602
Qy 559 WCGSGKLICTTAVPWNASWSNKSLEQIWNNTWMEWEREIDNYTNLIYTLIEESQNQOEK 618
Db 603 WCGSGKLICTTAVPWNASWSNKSLEQIWNNTWMEWDREINNYTSLIHSLEESQNQOEK 662
Qy 619 NEQELLELDKWASLWNWFDISKWLWYIK 646
Db 663 NEQELLELDKWASLWNWFNITNLWLWYIK 690

Search completed: February 25, 2004, 14:20:10
Job time : 20.8949 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:18:41 ; Search time 33.7043 Seconds
(without alignments)
4047.104 Million cell updates/sec

Title: US-09-891-609A-2
Perfect score: 3474
Sequence: 1 MRVKGIRKKNYQHLWRGGTLL.....DKWASLWNWFDISKWLWYIK 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3474	100.0	646	9 US-09-891-609-2	Sequence 2, Appli
2	3422	98.5	847	9 US-09-476-242-2	Sequence 2, Appli
3	3327	95.8	619	9 US-09-891-609-4	Sequence 4, Appli
4	3279.5	94.4	842	14 US-10-190-435-2	Sequence 2, Appli
5	3279.5	94.4	842	14 US-10-241-009-2	Sequence 2, Appli
6	3279.5	94.4	842	14 US-10-190-434B-2	Sequence 2, Appli
7	3279.5	94.4	842	14 US-10-190-305A-2	Sequence 2, Appli
8	2937.5	84.6	868	9 US-09-938-406-1	Sequence 1, Appli
9	2933	84.4	861	14 US-10-026-741-103	Sequence 103, App
10	2922.5	84.1	856	9 US-09-476-242-1	Sequence 1, Appli
11	2907.5	83.7	856	14 US-10-196-515-11	Sequence 11, Appl
12	2886	83.1	726	14 US-10-196-515-3	Sequence 3, Appli
13	2882	83.0	759	14 US-10-196-515-12	Sequence 12, Appl
14	2862	82.4	643	14 US-10-032-162-13	Sequence 13, Appl
15	2771	79.8	579	14 US-10-032-162-15	Sequence 15, Appl

16	2758	79.4	625	14	US-10-032-162-17	Sequence 17, Appl
17	2726	78.5	853	13	US-10-003-035-33	Sequence 33, Appl
18	2726	78.5	853	14	US-10-286-332A-33	Sequence 33, Appl
19	2726	78.5	853	15	US-10-280-915-33	Sequence 33, Appl
20	2700	77.7	860	14	US-10-190-435-6	Sequence 6, Appli
21	2700	77.7	860	14	US-10-241-009-6	Sequence 6, Appli
22	2700	77.7	860	14	US-10-190-434B-6	Sequence 6, Appli
23	2700	77.7	860	14	US-10-190-305A-6	Sequence 6, Appli
24	2667	76.8	870	14	US-10-190-435-147	Sequence 147, App
25	2642.5	76.1	867	14	US-10-190-435-3	Sequence 3, Appli
26	2642.5	76.1	867	14	US-10-190-435-126	Sequence 126, App
27	2642.5	76.1	867	14	US-10-241-009-3	Sequence 3, Appli
28	2642.5	76.1	867	14	US-10-190-434B-3	Sequence 3, Appli
29	2642.5	76.1	867	14	US-10-190-305A-3	Sequence 3, Appli
30	2640.5	76.0	865	14	US-10-190-435-140	Sequence 140, App
31	2635.5	75.9	855	14	US-10-190-435-144	Sequence 144, App
32	2634	75.8	861	14	US-10-190-435-139	Sequence 139, App
33	2632	75.8	858	14	US-10-190-435-150	Sequence 150, App
34	2628	75.6	862	14	US-10-190-435-141	Sequence 141, App
35	2626.5	75.6	855	14	US-10-369-294-11	Sequence 11, Appl
36	2626.5	75.6	855	14	US-10-369-294-12	Sequence 12, Appl
37	2624	75.5	870	14	US-10-190-435-127	Sequence 127, App
38	2621.5	75.5	869	14	US-10-190-435-4	Sequence 4, Appli
39	2621.5	75.5	869	14	US-10-241-009-4	Sequence 4, Appli
40	2621.5	75.5	869	14	US-10-190-434B-4	Sequence 4, Appli
41	2621.5	75.5	869	14	US-10-190-305A-4	Sequence 4, Appli
42	2619.5	75.4	845	14	US-10-190-435-129	Sequence 129, App
43	2619.5	75.4	1101	13	US-10-003-035-53	Sequence 53, Appl
44	2619.5	75.4	1101	14	US-10-286-332A-53	Sequence 53, Appl
45	2619.5	75.4	1101	15	US-10-280-915-53	Sequence 53, Appl

ALIGNMENTS

RESULT 1
US-09-891-609-2
; Sequence 2, Application US/09891609
; Patent No. US20020127238A1
; GENERAL INFORMATION:
; APPLICANT: Stamatatos, Leonidas
; TITLE OF INVENTION: HIV-1 VACCINES AND SCREENING METHODS THEREFOR
; FILE REFERENCE: 2570-1-002N
; CURRENT APPLICATION NUMBER: US/09/891,609
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,608
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-891-609-2

Query Match		100.0%;	Score 3474;	DB 9;	Length 646;
Best Local Similarity		100.0%;	Pred. No. 3.5e-310;		
Matches 646;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MRVKGIRKKNYQHLWRGGTLLGLMLICSAVEKLWVTVYGVVPVWKEATTTLFCASDAKAY	60		
Db	1	MRVKGIRKKNYQHLWRGGTLLGLMLICSAVEKLWVTVYGVVPVWKEATTTLFCASDAKAY	60		
Qy	61	DTEVHNWATHACVPTDPNPQEI VLENVTENFNMKNMVEQMHEDIISLWDQSLKPCVK	120		
Db	61	DTEVHNWATHACVPTDPNPQEI VLENVTENFNMKNMVEQMHEDIISLWDQSLKPCVK	120		
Qy	121	LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEKVGAGKLCNCNTSVITQACPKVS	180		
Db	121	LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEKVGAGKLCNCNTSVITQACPKVS	180		
Qy	181	FEPIPIHYCAPAGFAILKCNDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG	240		
Db	181	FEPIPIHYCAPAGFAILKCNDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG	240		

Db 181 FEPIPIHYCAPAGFAILKCNCKKFNKSGBCTNVSTVQCTHGIRPVVSTQLLLLNGSLAEEG 240
QY 241 VVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGAFYATGDIIGDIRQAH 300
Db 241 VVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGAFYATGDIIGDIRQAH 300
QY 301 CNISGEKWNNTLQIIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 360
Db 301 CNISGEKWNNTLQIIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 360
QY 361 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 420
Db 361 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 420
QY 421 GGKEISNTTEIFRPGGGMRDNRWSELYKYKVVVIEPLGVAPTAKRRVVQREKRAVTLG 480
Db 421 GGKEISNTTEIFRPGGGMRDNRWSELYKYKVVVIEPLGVAPTAKRRVVQREKRAVTLG 480
QY 481 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKLQ 540
Db 481 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKLQ 540
QY 541 ARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREIDN 600
Db 541 ARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREIDN 600
QY 601 YTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 646
Db 601 YTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 646

RESULT 2

US-09-476-242-2
; Sequence 2, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 2
; LENGTH: 847
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-476-242-2

Query Match 98.5%; Score 3422; DB 9; Length 847;
Best Local Similarity 95.4%; Pred. No. 3.1e-305;
Matches 643; Conservative 0; Mismatches 3; Indels 28; Gaps 2;
QY 1 MRVKGIRKNVQHLWRGGTLLGLMLMICS AVEKLVTVVYGVVPVWKEATTLFCASDAKAY 60
Db 1 MRVKGIRKNVQHLWRGGTLLGLMLMICS AVEKLVTVVYGVVPVWKEATTLFCASDAKAY 60
QY 61 DTEVHNWVWATHACVPTDPNPQEIIVLENVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWVWATHACVPTDPNPQEIIVLENVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 120
QY 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVTTISIRNMQKEYALFYKLDVV 180
QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKSGBCTNV 213
Db 181 PIDNDNTSYKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKSGBCTNV 240
QY 214 STVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNN 273

Db 241 STVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNN 300
QY 274 TRKSITIGPGAFYATGDIIGDIRQAHCNISGEKWNNTLQIIVTKLQAQFGNKTIVFKQS 333
Db 301 TRKSITIGPGAFYATGDIIGDIRQAHCNISGEKWNNTLQIIVTKLQAQFGNKTIVFKQS 360
QY 334 SGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNNTGTITLPCRIKQIINRWQEV 393
Db 361 SGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNNTGTITLPCRIKQIINRWQEV 420
QY 394 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGMRDNRWSELYKYKVV 453
Db 421 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGMRDNRWSELYKYKVV 480
QY 454 KIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIV 513
Db 481 KIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIV 540
QY 514 QQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGWCSGKLICTTAVP 572
Db 541 QQNNLLRAIEAQHLLQLTVWGIKQARVLAVERYLKDQQLLGWCSGKLICTTAVP 600
QY 573 WNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASL 632
Db 601 WNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASL 660
QY 633 WNWFDISKWLWYIK 646
Db 661 WNWFDISKWLWYIK 674

RESULT 3

US-09-891-609-4
; Sequence 4, Application US/09891609
; Patent No. US20020127238A1
; GENERAL INFORMATION:
; APPLICANT: Stamatatos, Leonidas
; TITLE OF INVENTION: HIV-1 VACCINES AND SCREENING METHODS THEREFOR
; FILE REFERENCE: 2570-1-002N
; CURRENT APPLICATION NUMBER: US/09/891,609
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,608
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-891-609-4

Query Match 95.8%; Score 3327; DB 9; Length 619;
Best Local Similarity 100.0%; Pred. No. 1.1e-296;
Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 SAVEKLVWTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 87
Db 1 SAVEKLVWTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60
QY 88 VTENFNMMKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 147
Db 61 VTENFNMMKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
QY 148 DRGEIKNCSFKVGAGKLIINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKS 207
Db 121 DRGEIKNCSFKVGAGKLIINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKS 180
QY 208 GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINC 267
Db 181 GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINC 240
QY 268 TRPNNNTRKSIITIGPGAFYATGDIIGDIRQAHCNISGEKWNNTLQIIVTKLQAQFGNKT 327

Db 241 TRPNNTRKSIITIGPGAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKT 300
Qy 328 IVFKQSSGGDPPIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNNTGTITLPCRICKII 387
Db 301 IVFKQSSGGDPPIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNNTGTITLPCRICKII 360
Qy 388 NRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 447
Db 361 NRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 420
Qy 448 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAGSTMGARSLTLTVQARQ 507
Db 421 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAGSTMGARSLTLTVQARQ 480
Qy 508 LLSGIVQQNNLLRAIEAQOHLQLTVMGIKQARVLAVERYLKDQQLLGIWCSGKLIC 567
Db 481 LLSGIVQQNNLLRAIEAQOHLQLTVMGIKQARVLAVERYLKDQQLLGIWCSGKLIC 540
Qy 568 TTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELLELD 627
Db 541 TTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELLELD 600
Qy 628 KWASLWNWFDISKWLWYIK 646
Db 601 KWASLWNWFDISKWLWYIK 619

RESULT 4
US-10-190-435-2
; Sequence 2, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PPI8133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SF162

Query Match 94.4%; Score 3279.5; DB 14; Length 842;
Best Local Similarity 94.2%; Pred. No. 3.9e-292;
Matches 619; Conservative 2; Mismatches 7; Indels 29; Gaps 3;
Qy 19 LLLGMLMIC-SAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 77
Db 13 LLCGAVFVSPSAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 72
Qy 78 PNPQEIIVLENTNFNMWKNMVEQMHEDIISLDQSLKPCVKLTPLCVTLHCTNLKNAT 137
Db 73 PNPQEIIVLENTNFNMWKNMVEQMHEDIISLDQSLKPCVKLTPLCVTLHCTNLKNAT 132
Qy 138 NTKSSNWKEMDRGEIKNCSEFKVGAG-----KLINCNNTS 170
Db 133 NTKSSNWKEMDRGEIKNCSEFKVTTSIRNMQKEYALFYKLDVVPIDNDNTSYKLINCNTS 192
Qy 171 VITQACPKVSFEPIPIHYCAPAGFAILKCNDDKFNKFGSGPCTNVSTVQCTHGIRPVVSTQL 230
Db 193 VITQACPKVSFEPIPIHYCAPAGFAILKCNDDKFNKFGSGPCTNVSTVQCTHGIRPVVSTQL 252
Qy 231 LLNGSLAEEGVIRSENFDTNNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGAFYATG 290

Db 253 LLNGSLAEEGVIRSENFDTNNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGAFYATG 312
Qy 291 DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPPIVMHSFNCGG 350
Db 313 DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPPIVMHSFNCGG 372
Qy 351 FFYCNSTQLFNSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSS 410
Db 373 FFYCNSTQLFNSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSS 432
Qy 411 NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVVIEPLGVAPTAKRRVV 470
Db 433 NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVVIEPLGVAPTAKRRVV 492
Qy 471 QREKRAVTLGAMFLGLGAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL 530
Db 493 QREKRAVTLGAMFLGLGAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL 552
Qy 531 QLTVMGIK-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLDQIWN 589
Db 553 QLTVMGIKQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLDQIWN 612
Qy 590 TWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 646
Db 613 TWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 5
US-10-241-009-2
; Sequence 2, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SF162

Query Match 94.4%; Score 3279.5; DB 14; Length 842;
Best Local Similarity 94.2%; Pred. No. 3.9e-292;
Matches 619; Conservative 2; Mismatches 7; Indels 29; Gaps 3;
Qy 19 LLLGMLMIC-SAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 77
Db 13 LLCGAVFVSPSAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 72
Qy 78 PNPQEIIVLENTNFNMWKNMVEQMHEDIISLDQSLKPCVKLTPLCVTLHCTNLKNAT 137
Db 73 PNPQEIIVLENTNFNMWKNMVEQMHEDIISLDQSLKPCVKLTPLCVTLHCTNLKNAT 132
Qy 138 NTKSSNWKEMDRGEIKNCSEFKVGAG-----KLINCNNTS 170
Db 133 NTKSSNWKEMDRGEIKNCSEFKVTTSIRNMQKEYALFYKLDVVPIDNDNTSYKLINCNTS 192
Qy 171 VITQACPKVSFEPIPIHYCAPAGFAILKCNDDKFNKFGSGPCTNVSTVQCTHGIRPVVSTQL 230
Db 193 VITQACPKVSFEPIPIHYCAPAGFAILKCNDDKFNKFGSGPCTNVSTVQCTHGIRPVVSTQL 252
Qy 231 LLNGSLAEEGVIRSENFDTNNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGAFYATG 290

Db 253 LLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATG 312

Qy 291 DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKOSSGGDPEIVMHSFNCGGE 350

Db 313 DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKOSSGGDPEIVMHSFNCGGE 372

Qy 351 FFYCNSTQLFNSWTWNNTIGPNNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS 410

Db 373 FFYCNSTQLFNSWTWNNTIGPNNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS 432

Qy 411 NITGLLLTRDGGKEISNTTEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKARRVV 470

Db 433 NITGLLLTRDGGKEISNTTEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKARRVV 492

Qy 471 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 530

Db 493 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 552

Qy 531 QLTVMGIK-LQARVLAVERYLKQQLLGIWCGSGKLICTTAVPWNASWSNKSOLDQIWNMM 589

Db 553 QLTVMGIKQLOARVLAVERYLKQQLLGIWCGSGKLICTTAVPWNASWSNKSOLDQIWNMM 612

Qy 590 TWMEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 646

Db 613 TWMEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 6

US-10-190-434B-2

; Sequence 2, Application US/10190434B

; Publication No. US20030194800A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MEGEDE, Jan

; APPLICANT: BARNETT, Susan

; APPLICANT: LIAN, Ying

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B

; FILE REFERENCE: 2300-1621.20

; CURRENT APPLICATION NUMBER: US/10/190,434B

; CURRENT FILING DATE: 2002-07-05

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 842

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SF162

US-10-190-434B-2

Query Match 94.4%; Score 3279.5; DB 14; Length 842;

Best Local Similarity 94.2%; Pred. No. 3.9e-292;

Matches 619; Conservative 2; Mismatches 7; Indels 29; Gaps 3;

Qy 19 LLLGMLMIC-SAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 77

Db 13 LLCGAVFVSPSAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 72

Qy 78 PNPQEIIVLENTVENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT 137

Db 73 PNPQEIIVLENTVENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT 132

Qy 138 NTKSSNWKEMDRGEIKNCSEFKVGAG-----KLINCNTS 170

Db 133 NTKSSNWKEMDRGEIKNCSEFKVTTTSIRNMKQKEYALFYKLDVVPIDNDNTSYKLINCNTS 192

Qy 171 VITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQL 230

Db 193 VITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQL 252

Qy 231 LLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATG 290

Db 253 LLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATG 312

Qy 291 DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKOSSGGDPEIVMHSFNCGGE 350

Db 313 DIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKOSSGGDPEIVMHSFNCGGE 372

Qy 351 FFYCNSTQLFNSWTWNNTIGPNNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS 410

Db 373 FFYCNSTQLFNSWTWNNTIGPNNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS 432

Qy 411 NITGLLLTRDGGKEISNTTEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKARRVV 470

Db 433 NITGLLLTRDGGKEISNTTEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKARRVV 492

Qy 471 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 530

Db 493 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 552

Qy 531 QLTVMGIK-LQARVLAVERYLKQQLLGIWCGSGKLICTTAVPWNASWSNKSOLDQIWNMM 589

Db 553 QLTVMGIKQLOARVLAVERYLKQQLLGIWCGSGKLICTTAVPWNASWSNKSOLDQIWNMM 612

Qy 590 TWMEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 646

Db 613 TWMEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 7

US-10-190-305A-2

; Sequence 2, Application US/10190305A

; Publication No. US20030198621A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MEGEDE, Jan

; APPLICANT: BARNETT, Susan

; APPLICANT: LIAN, Ying

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR

; FILE REFERENCE: 2302-18702 / 18702.002

; CURRENT APPLICATION NUMBER: US/10/190,305A

; CURRENT FILING DATE: 2002-07-05

; NUMBER OF SEQ ID NOS: 93

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 842

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SF162

US-10-190-305A-2

Query Match 94.4%; Score 3279.5; DB 14; Length 842;

Best Local Similarity 94.2%; Pred. No. 3.9e-292;

Matches 619; Conservative 2; Mismatches 7; Indels 29; Gaps 3;

Qy 19 LLLGMLMIC-SAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 77

Db 13 LLCGAVFVSPSAVEKLVTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTD 72

Qy 78 PNPQEIIVLENTVENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT 137

Db 73 PNPQEIIVLENTVENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAT 132

Qy 138 NTKSSNWKEMDRGEIKNCSEFKVGAG-----KLINCNTS 170

Db 133 NTKSSNWKEMDRGEIKNCSEFKVTTTSIRNMKQKEYALFYKLDVVPIDNDNTSYKLINCNTS 192

Qy 171 VITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQL 230

Db 193 VITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQL 252

Qy 231 LLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATG 290

Db 253 LLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATG 312

QY 291 DIIGDIRQAHCHNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPPIVHSHFNCGGE 350
Db 313 DIIGDIRQAHCHNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPPIVHSHFNCGGE 372
QY 351 FFYCNSTQLFNSHTWNTTIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS 410
Db 373 FFYCNSTQLFNSHTWNTTIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS 432
QY 411 NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVVIEPLGVAPTAKRRVV 470
Db 433 NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVVIEPLGVAPTAKRRVV 492
QY 471 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 530
Db 493 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 552
QY 531 QLTVMGIK-LQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLLDQIWNMM 589
Db 553 QLTVMGIKQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLLDQIWNMM 612
QY 590 TWMEWEREIDNYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWFDISKWLWYIK 646
Db 613 TWMEWEREIDNYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 8
US-09-938-406-1
; Sequence 1, Application US/09938406
; Patent No. US20020155120A1
; GENERAL INFORMATION:
; APPLICANT: Lowell, George
; APPLICANT: Vancott, Thomas
; APPLICANT: Birx, Deborah
; TITLE OF INVENTION: PROTEIN AND PEPTIDE VACCINES FOR
; FILE REFERENCE: 40646-20002.10
; CURRENT APPLICATION NUMBER: US/09/938,406
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 09/214,701
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: PCT/US 97/12253
; PRIOR FILING DATE: 1997-07-10
; PRIOR APPLICATION NUMBER: US 60/021,687
; PRIOR FILING DATE: 1996-07-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Virus HIV-1
US-09-938-406-1

Query Match 84.6%; Score 2937.5; DB 9; Length 868;
Best Local Similarity 79.5%; Pred. No. 1.1e-260;
Matches 556; Conservative 37; Mismatches 47; Indels 59; Gaps 8;

QY 1 MRVKGIRKNYQHLWRGGTLLGLMLMICSAAVEKLVTVYGVVPVWKEATTLFCASDAKAY 60
Db 3 MRAKGIRKNQHLWRWGTMLLGLMLMICSAAANLVTVYGVVPVWKEATTLFCASDAKAY 62
QY 61 DTEVHNWVATHACVPTDPNPQEIIVLENVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 120
Db 63 DTEAHNVWATHACVPTNPQEVVLENVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 122
QY 121 LTPLCVTLHCT--NLKNATNTKSSN---WKEMDRGEIKNCSFKVGAG----- 162
Db 123 LTPLCVTLNCTDLNTNTTNTTTELSIIVVWEQKGEMRNCNFITTSIRDKVQREYALF 182
QY 163 -----KLNCSNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCK 202
Db 183 YKLDVEPIDDNKNTNTKYRLINCNTSVITQACPKVSFEPIPIHYCTPTGFALLKCNCK 242
QY 203 KFNCGSPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAGVVIRSENFTDNAKTIIVQLKES 262

Db 243 KFNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAGVVIRSENFTDNAKTIIVQLNVS 302
QY 263 VEINCTRPNNNTRKSIITIGPRAFYATGDIIGDIRQAHCHNISGEKWNNTLKQIVTKLQAO 322
Db 303 VEINCTRPNNNTRKRVTLGPGRVYTTGEILGNIRQAHCHNISRAQWNNTLQOIATTLREQ 362
QY 323 FGNKTIVFKQSSGGDPPIVHSHFNCGGEFFYCNSTQLFNSHTWNTTIGPNTNGT----- 376
Db 363 FGNKTIAFNQSSGGDPPIVHSHFNCGGEFFYCNSTQLFNSAWNVT-----SNGTWSVTRK 417
QY 377 -----ITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTT 429
Db 418 QKTDGDIITLPCRIKQIINRWQEVGKAMYALPIKGLIRCSSNITGLLLTRDGGGE-NQTT 476
QY 430 EIFRPGGDMRDNRSELYKYKVVVIEPLGVAPTAKRRVVQREKRAV-TLGAMFLGFLG 488
Db 477 EIFRPGGDMRDNRSELYKYKVVVIEPLGVAPTAKRRVVQREKRAVGMGLGAMFLGFLG 536
QY 489 AAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVMGIK-LQARVLAVE 547
Db 537 AAGSTMGATSMALTQARQLLSGIVQQNNLLRAIKAQQHLLQLTVMGIKQLQARILAVE 596
QY 548 RYLKDOQLLGIWCGSGKLICTTAVPWNASWSNKSLLDQIWNMTWMEWEREIDNYTNLIYT 607
Db 597 RYLKDOQLLGFWCGSGKLICTTAVPWNASWSNKTLDQIWNMTWMEWDREIDNYTHLIYT 656
QY 608 LIEESONQOEKNEQELLELDKWASLWNWFDISKWLWYIK 646
Db 657 LIEESONQOEKNOQELLELDKWASLWTSWSDITKWLWYIK 695

RESULT 9
US-10-026-741-103
; Sequence 103, Application US/10026741
; Publication No. US20030049604A1
; GENERAL INFORMATION:
; APPLICANT: CHARNEAU, PIERRE
; CLAVEL, FRANCOISE
; BORMAN, ANDREW
; QUILLET, CAROLINE
; GUETARD, DENISE
; MONTAGNIER, LUC
; DONJON DE SAINT-MARTIN, JACQUELINE
; COHEN, JACQUES
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR
; SUBTYPE) ANTIGENS
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/026,741
; FILING DATE: 27-Dec-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/817,441
; FILING DATE: 31-AUG-1998
; APPLICATION NUMBER: PCT/FR 95/01391
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: FR 9412554
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: FR 9502526
; FILING DATE: 03-MAR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03260.6005-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 861 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-10-026-741-103

```
Query Match      84.4%; Score 2933; DB 14; Length 861;
Best Local Similarity 81.8%; Pred. No. 2.8e-260;
Matches 567; Conservative 33; Mismatches 41; Indels 52; Gaps 12;

QY 1 MRVKGIRKNYQHLWR---GGTLLGLMLMICSAVEKLWTVVYGVVWKEATTTLFCASD 56
   ||||| : ||||| ||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MRVK---EKYQHLWRGWKGWGTLLGILMICSAATEKLWTVVYGVVWKEATTTLFCASD 57

QY 57 AKAYDTEVHNWATHACVPTDPNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLK 116
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 AKAYDTEVHNWATHACVPTDPNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLK 117

QY 117 PCVKLTPLCVTLHCTNLKNATNTKSSNWKE-----MDRGEIKNCSFKVGAG---GK--- 163
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 PCVKLTPLCVSLKCTDLGNATNTNSSNTSSSGEMMEKEIKNCSFNISTSRGKVQKE 177

QY 164 -----LINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKK 203
   ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 YAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKT 237

QY 204 FNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFNTDNAKTIIVQLKESV 263
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 FNGTGPCNTVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFNTDNAKTIIVQLNQSV 297

QY 264 EINCTRPNNTKRSITI--GPGRAFATGDIIGDIRQAHNCNISGEKWNNTLKQIVTKLOA 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 EINCTRPNNTKRSIRIQRGPGRAFVTIGK-IGNMRQAHNCNISRAKWNATLKQIASKLRE 356

QY 322 QFG-NKTIIVFKOSSGGDPEIIVMHSFNCGGFFFCNSTQLFNSTWNN---TIGPNTNG- 375
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 QFGNNKTIIFKOSSGGDPEIIVTHSFNCGGFFFCNSTQLFNSTWFSNTEGSNNTSGS 416

QY 376 -TITLPCRIOIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRP 434
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 DTITLPCRIOFINMWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNN-NNGSEIFRP 475

QY 435 GGGMRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAAGSTM 494
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 GGGMRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGVIGALFLGLGAAGSTM 535

QY 495 GARSLLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKQD 553
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 536 GARSMTLTQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQARILAVERYLKQD 595

QY 554 QLLGIWCGSKLICTTAVPWNASWSNKSLDQIWNNTMTWEMEREIDNYTNLIYTLIEESQ 613
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 596 QLLGIWCGSKLICTTAVPWNASWSNKSLEQIWNNTMTWEMEDREINNYSLSLHSLIEESQ 655

QY 614 NQEKNEQELLELDKWASLWNWFDISKWLWYIK 646
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 656 NQEKNEQELLELDKWASLWNWFDITNWLWYIK 688
```

RESULT 10
US-09-476-242-1
; Sequence 1, Application US/09476242

Patent No. US20020146683A1
GENERAL INFORMATION:
APPLICANT: BARNETT, Susan
APPLICANT: HARTOG, Karin
APPLICANT: MARTIN, Eric
TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
FILE REFERENCE: 1605.002
CURRENT APPLICATION NUMBER: US/09/476,242
CURRENT FILING DATE: 1999-12-30
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 856
TYPE: PRT
ORGANISM: Human immunodeficiency virus
US-09-476-242-1

```
Query Match      84.1%; Score 2922.5; DB 9; Length 856;
Best Local Similarity 81.8%; Pred. No. 2.6e-259;
Matches 563; Conservative 35; Mismatches 43; Indels 47; Gaps 11;

QY 1 MRVKGIRKNYQHL---WRGGTLLGLMLMICSAVEKLWTVVYGVVWKEATTTLFCASD 56
   ||||| : ||||| ||: ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MRVK---EKYQHLWRGWGTLLGMLMICSAATEKLWTVVYGVVWKEATTTLFCASD 57

QY 57 AKAYDTEVHNWATHACVPTDPNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLK 116
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 AKAYDTEVHNWATHACVPTDPNPQEIIVLENTNFNMWKNMVEQMHEDIISLWDQSLK 117

QY 117 PCVKLTPLCVTLHCTNLKNATNTKSSNWKE-MDRGEIKNCSFKVGAG----- 162
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 PCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKEIKNCSFNISTSRGKVQKEYAFFY 177

QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSG 208
   ||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 KLDIIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKT FNGTG 237

QY 209 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFNTDNAKTIIVQLKESVEINCT 268
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 238 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFNTDNAKTIIVQLNTEVEINCT 297

QY 269 RPNNTKRSITI--GPGRAFATGDIIGDIRQAHNCNISGEKWNNTLKQIVTKLOAQFG-N 325
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 298 RPNNTKRSIRIQRGPGRAFVTIGK-IGNMRQAHNCNISRAKWNNTLKQIASKLREQFGNN 356

QY 326 KTIIVFKOSSGGDPEIIVMHSFNCGGFFFCNSTQLFNSTWNN---TIGPNTNG--TITL 379
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 KTIIFKOSSGGDPEIIVTHSFNCGGFFFCNSTQLFNSTWFSNTEGSNNTSGSDTITL 416

QY 380 PCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDM 439
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 PCRIKQIINMWQKVGKAMYAPPISGQIRCSSNITGLLLTRDGGNS-NNESEIFRPGGDM 475

QY 440 RDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAAGSTMGARSL 499
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 RDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGVIGALFLGLGAAGSTMGAASM 535

QY 500 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGI 558
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 536 TLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQARILAVERYLKDQQLLGI 595

QY 559 WCGSKLICTTAVPWNASWSNKSLDQIWNNTMTWEMEREIDNYTNLIYTLIEESQNQOEK 618
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 596 WCGSKLICTTAVPWNASWSNKSLEQIWNHTTWEMEDREINNYSLSLHSLIEESQNQOEK 655

QY 619 NEQELLELDKWASLWNWFDISKWLWYIK 646
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 656 NEQELLELDKWASLWNWFDITNWLWYIK 683
```

RESULT 11
US-10-196-515-11
; Sequence 11, Application US/10196515


```
; Publication No. US20030091594A1
; GENERAL INFORMATION:
; APPLICANT: HOXIE, James A.
; APPLICANT: LABRANCHE, Celia C.
; APPLICANT: DOMS, Robert W.
; APPLICANT: HOFFMAN, Trevor L.
; TITLE OF INVENTION: CD4-INDEPENDENT HIV ENVELOPE PROTEINS AS VACCINES AND THERAPEUTICS
; FILE REFERENCE: Hoxie 9596-104U1 (0282)
; CURRENT APPLICATION NUMBER: US/10/196,515
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/337,387
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: US 09/317,556
; PRIOR FILING DATE: 1999-05-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-196-515-11

Query Match      83.7%; Score 2907.5; DB 14; Length 856;
Best Local Similarity 81.7%; Pred. No. 6.2e-258;
Matches 562; Conservative 37; Mismatches 42; Indels 47; Gaps 12;

QY   1 MRVKGIRKNYQHL-----WRGGTLLGLMLMICA VEKLVWTVVYGVPVWKEATTTLFCASD 56
    : ||||| |:::|||||:|||||||:|||||||:|||||||:|||||||:|||||||:
Db   1 MRVK--EKYQHLLRWGRWGTMLLGLMLMICNATEKLVWTVVYGVPVWKEATTTLFCASD 57

QY   57 AKAYDTEVHNWATHACVPTDPNPQEIVLENTENFNWKNNMVQMHEDIISLDQSILK 116
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   58 AKAYDTEVHNWATHACVPTDPNPQEVVLNVNTENFOMWKNDMVQMHEDIISLDQSILK 117

QY   117 PCVKLTPLCVTLHCTNLKNATNTKSSNWKE-MDRGEIKNCSPFKVGA---GK----- 163
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   118 PCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKEIKNCSFNISTSI RGVKEYAFY 177

QY   164 -----LINCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNDKKFNGSG 208
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   178 KLDIIPIDNDTTSYLSCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNDKTFTGTG 237

QY   209 PCTNVSTVQCTHGIRPVYSTQLLLNGSLAEAGVVIRSENFTDNAKTIIVQLKESVEINCT 268
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   238 PCTNVSTVQCTHGIRPVYSTQLLLNGSLAEAEVVIRSVNFTDNAKTIIVQLNTSVSEINCT 297

QY   269 RPNNTNRKSITI--GPGRFYATGDIIGDIRQAHCNISGEKWNTLKQIVTKLAQFG-N 325
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   298 RPNNTNRKRIRIQRGPGRAFVTIGK-IGNMQRAHCNISRANKWNTLKQIDSKLRQFGNN 356

QY   326 KTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNV----TGPNNTNG--TITL 379
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   357 KTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSNTEGSDTITL 416

QY   380 PCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLTRDDGGKEISNTTEIFRPGGDM 439
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   417 PCRICKQIINMWQKVGMKYAPPISGQIRCSSNITGLLTRDDGGNS--NNESEIFRPGGDM 475

QY   440 RDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSL 499
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   476 RDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASM 535

QY   500 TLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGVIK-LQARVLAVERYLKDOQLLGI 558
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   536 TLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGVIKQLQARILAVERYLKDOQLLGI 595

QY   559 WGCSGKLICTTAVPNAWSNKS LDQIWNMTWMWEEREIDNYTNLIYTLEE SQNQEK 618
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db   596 WGCSGKLICTTAVPNAWSNKS LEQIWNHTTWMEWDREINNYTSLIHSLIE SQNQEK 655

QY   619 NEQELLELDK WASLNWFDISKWLWYIK 646
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


Db 420 GGIN-ENGTEIFRPGGDMRDNRSEFYKYVVKIEPLGVAPTCKRRVVQREKRAVGIG 478
QY 481 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGIK-L 539
Db 479 AVFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQNNLLRAIEAQOQMLQLTVWGIKQL 538
QY 540 QARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLDQIWNNTWMEWEREID 599
Db 539 QARVLAVERYLGDQQLLGIWCGSGKLICTTAVPWNASWSNKSLDRIWNNTWMEWEREID 598
QY 600 NYTNLIYTLIEESQNQEKNEQEELLELDKWASLWNWFDISKWLWY 644
Db 599 NYTSEIYTLIEESQNQEKNEQEELLELDKWASLWNWFDITNWLWY 643

RESULT 15
US-10-032-162-15
; Sequence 15, Application US/10032162
; Publication No. US20030052839A1
; GENERAL INFORMATION:
; APPLICANT: BINLEY, JAMES M
; APPLICANT: SCHUELKE, NORBERT
; APPLICANT: OLSON, WILLIAM C
; APPLICANT: PAUL, MADDON J
; APPLICANT: JOHN, MOORE P
; TITLE OF INVENTION: STABILIZED VIRAL ENVELOPE PROTEINS AND USES THEREOF
; FILE REFERENCE: 2048/59331az
; CURRENT APPLICATION NUMBER: US/10/032,162
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 09/602,864
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-032-162-15

Query Match 79.8%; Score 2771; DB 14; Length 579;
Best Local Similarity 85.6%; Pred. No. 1.2e-245;
Matches 529; Conservative 16; Mismatches 31; Indels 42; Gaps 6;
QY 30 VEKLWTVVYGVVPWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLENTV 89
Db 1 VEKLWTVVYGVVPWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVLENTV 60
QY 90 ENFNWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEMDR 149
Db 61 EHFNMWKNMVEQMQEDIISLWDSLKPCVKLTPLC----- 96
QY 150 GEIKNCSFKVGAGKLLINCNTSVITQACPKVSFEPPIHYCAPAGFAILKCNCKKFNCGSP 209
Db 97 -----GAG-----CDTSVITQACPKISFEPPIHYCAPAGFAILKCNCKTFNGKGP 142
QY 210 CTNVSTVQCTHGIRPVVSTQLLNGSLAEEGWIRSENFTDNAKTIIVQLKESVEINCTR 269
Db 143 CKNVSTVQCTHGIRPVVSTQLLNGSLAEFEWIRSDNFTNNAKTIIVQLKESVEINCTR 202
QY 270 PNNTRKSTIGPGRAFATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIV 329
Db 203 PNNTRKSIHIGPGRAFYTGTGEIIGDIRQAHCNISRAKWNNTLKQIVIKLREQFENKTIV 262
QY 330 FKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTW-NNTIGNNTNG-TITLPCRICKI 387
Db 263 FNHSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTEGNNTEGNTITLPCRICKI 322
QY 388 NRWQEVGKAMYAPPPIRGQIRCSSNITGILLTRDGGKEISNTTEIFRPGGDMRDNRSEL 447
Db 323 NMWQEVGKAMYAPPPIRGQIRCSSNITGILLTRDGGIN-ENGTEIFRPGGDMRDNRSEL 381
QY 448 YKYKVVKIEPLGVAPTCKRRVVQREKRAVTLGAMFLGLGAAGSTMGARSLTLTVQARQ 507

Db 382 YKYKVVKIEPLGVAPTCKRRVVQREKRAVGIGAVFLGLGAAGSTMGAASMTLTVQARL 441
QY 508 LLSGIVQQNNLLRAIEAQOHLQLTVWGIK-LQARVLAVERYLKDQQLLGIWCGSGKLI 566
Db 442 LLSGIVQQNNLLRAIEAQOQMLQLTVWGIKQLQARVLAVERYLGDQQLLGIWCGSGKLI 501
QY 567 CTTAVPWNASWSNKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQEKNEQEELLE 626
Db 502 CCTAVPWNASWSNKSLDRIWNNTWMEWEREIDNYTSEIYTLIEESQNQEKNEQEELLE 561
QY 627 DKWASLWNWFDISKWLWY 644
Db 562 DKWASLWNWFDITNWLWY 579

Search completed: February 25, 2004, 14:28:09
Job time : 36.7043 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: February 25, 2004, 14:12:55 ; Search time 18.3842 Seconds
(without alignments)
3380.063 Million cell updates/sec

Title: US-09-891-609A-2
Perfect score: 3474
Sequence: 1 MRVKGIRKNYQHLLWRGGTLL.....DKWASLWNWFDISKWLWYIK 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pirl:*
2: pirl2:*
3: pirl3:*
4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3035	87.4	847	2 T09448	envelope glycoprot
2	3032	87.3	843	1 H44001	env polyprotein pr
3	3025.5	87.1	852	2 T12016	envelope glycoprot
4	3008	86.6	847	2 S13289	env protein - huma
5	2984	85.9	855	1 VCLJA2	env polyprotein pr
6	2953	85.0	861	1 VCLJSC	env polyprotein pr
7	2939.5	84.6	856	1 VCLJH3	env polyprotein pr
8	2938.5	84.6	852	1 VCLJBR	env polyprotein -
9	2937.5	84.6	868	1 VCLJH4	env polyprotein pr
10	2933	84.4	861	1 VCLJLV	env polyprotein pr
11	2928	84.3	851	2 S33985	env polyprotein -
12	2926.5	84.2	856	1 VCLJVL	env polyprotein pr
13	2914.5	83.9	854	2 S13288	env protein - huma
14	2908.5	83.7	729	1 VCLJKX	env polyprotein pr
15	2908.5	83.7	861	1 VCLJKB	env polyprotein pr
16	2894.5	83.3	856	1 VCLJ3W	env polyprotein pr
17	2890.5	83.2	859	1 VCLJMN	env polyprotein pr
18	2678.5	77.1	859	2 T01672	envelope polyprote
19	2672	76.9	855	1 VCLJZR	env polyprotein pr
20	2663	76.7	853	2 S54384	envelope polyprote
21	2632.5	75.8	856	1 A44963	env polyprotein pr
22	2611.5	75.2	846	1 VCLJND	env polyprotein pr
23	2196	63.2	854	1 VCLJSI	env polyprotein pr
24	2090.5	60.2	506	2 A40218	envelop glycoprote
25	1945.5	56.0	445	2 A41621	env polyprotein M
26	1941.5	55.9	443	2 C41621	env polyprotein p
27	1849	53.2	454	2 B41621	env polyprotein D
28	1848	53.2	495	2 S31493	env polyprotein -
29	1830.5	52.7	877	2 S49197	envelope protein p

30	1750.5	50.4	863	2 A53034	gag polyprotein -
31	1301	37.4	290	2 S25940	env protein - huma
32	1286	37.0	297	2 S60538	envelope polyprote
33	1189	34.2	299	2 S60528	envelope polyprote
34	1188.5	34.2	294	2 S60545	envelope polyprote
35	1187.5	34.2	294	2 S60524	envelope polyprote
36	1186	34.1	299	2 S60529	envelope polyprote
37	1183	34.1	877	2 C46356	env polyprotein -
38	1179	33.9	301	2 S60548	envelope polyprote
39	1173.5	33.8	300	2 S60546	envelope polyprote
40	1169.5	33.7	300	2 S60547	envelope polyprote
41	1169	33.6	299	2 S60521	envelope polyprote
42	1167.5	33.6	852	1 VCLJGG	env polyprotein pr
43	1166	33.6	859	1 VCLJST	env polyprotein pr
44	1163	33.5	301	2 S60532	envelope polyprote
45	1163	33.5	301	2 S60531	envelope polyprote

ALIGNMENTS

RESULT 1

T09448

envelope glycoprotein - human immunodeficiency virus type 1 (strain JRFL)
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 24-Nov-1999
C:Accession: T09448

R;Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie, submitted to the EMBL Data Library, July 1996

A:Reference number: Z16673

A:Accession: T09448

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-847 <PAN>

A:Cross-references: EMBL:U63632; NID:gl465777; PID:gl465781

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

Query Match 87.4%; Score 3035; DB 2; Length 847;
Best Local Similarity 85.4%; Pred. No. 1.7e-211;
Matches 577; Conservative 29; Mismatches 38; Indels 32; Gaps 6;

Qy	1	MRVKGIRKNYQHLLWRGGTLLGLMLICSAVEKLWTVVYGVVWKEATTLFCASDAKAY	60
Db	1	MRVKGIRKSYQYLWKGGTLLGLMLICSAVEKLWTVVYGVVWKEATTLFCASDAKAY	60
Qy	61	DTEVHNWATHACVPTDPNPQEI VLENTENFNWKNMVEMQHEDIISLWDQSLKPCVK	120
Db	61	DTEVHNWATHACVPTDPNPQEVVLENTHEFNWKNMVEMQEDIISLWDQSLKPCVK	120
Qy	121	LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG-----	162
Db	121	LTPLCVTLNCKDV-NATNTTNDSEGTMERGEIKNCSFNITTSIRDEVQKEYALFYKLDVV	179
Qy	163	-----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNV	213
Db	180	PIDNNNTSYRLISCDTSVITQACPKISFEPIPIHYCAPAGFAILKCNDKTFNGKGPCKNV	239
Qy	214	STVQCTHGIRPVVSTQLLNGSLAEEGVWIRSENFTDNAKTIIVQLKESVEINCTRPNNN	273
Db	240	STVQCTHGIRPVVSTQLLNGSLAEEVWIRSDNFTNNAKTIIVQLKESVEINCTRPNNN	299
Qy	274	TRKSITIGPGRAFATGDIIGDIRQAHCNISGEKNNTLKQIVTKLQAOFGNKTIIVFKQS	333
Db	300	TRKSIHIGPGRAFYTGTGEIIGDIRQAHCNISRAKWNDTLKQIVIKLREQFENKTIIVFNHS	359
Qy	334	SGGDPEIVMHSFNCGGEFFYCNSTQLFNSTW-NNTIGPNNNTG-TITLPCRICKIINRWQ	391
Db	360	SGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNNTSGSNTEGNTITLPCRICKIINMWQ	419
Qy	392	EVGKAMVAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYK	451

Db 420 EVGKAMYAPPIRGQIRCSSNITGLLLTRDGGIN-ENGTEIFRPGGDMRDNRSELYKYK 478

Qy 452 VVKIEPLGVAPTAKARRVVQREKRAVTLGAMFLGFLGAAGSTMGAARSLTLTVQARQLLSG 511

Db 479 VVKIEPLGVAPTAKARRVVQREKRAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSG 538

Qy 512 IVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQLLGIWCSGKLICTTA 570

Db 539 IVQQNNLLRAIEAQORMLQLTVWGIKLQARVLAVERYLGDQLLGIWCSGKLICTTA 598

Qy 571 VPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWA 630

Db 599 VPWNASWSNKSLDRIWNMTWMEWEREIDNYTSEIYTLIEESNQOEKNEQELLELDKWA 658

Qy 631 SLWNWFDISKWLMYIK 646

Db 659 SLWNWFDITKWLWYIK 674

RESULT 2

H44001

env polyprotein precursor - human immunodeficiency virus type 1 (strain YU-2)

N;Alternate names: coat polyprotein

N;Contains: coat protein gp120; coat protein gp41

C;Species: human immunodeficiency virus type 1, HIV-1

A;Note: host Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Sep-1994

C;Accession: H44001

R;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.

J. Virol. 66, 6587-6600, 1992

A;Title: Complete nucleotide sequence, genome organization, and biological properties of

A;Reference number: A44001; MUID:93021387; PMID:1404605

A;Accession: H44001

A;Molecule type: DNA

A;Residues: 1-843 <LIY>

A;Cross-references: GB:M93258

C;Genetics:

A;Gene: env

C;Superfamily: type E retrovirus env polyprotein

C;Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein

F;1-29/Domain: signal sequence #status predicted <SIG>

F;19-35/Region: hydrophobic

F;30-489/Product: coat protein gp120 #status predicted <GP1>

F;490-843/Product: coat protein gp41 #status predicted <GP2>

F;499-515/Region: hydrophobic

F;673-689/Region: hydrophobic

F;738-755/Domain: transmembrane #status predicted <TMN>

F;87,129,135,138,154,158,184,193,230,237,258,272,285,291,297,327,351,381,389,395,400,435

Query Match 87.3%; Score 3032; DB 1; Length 843;

Best Local Similarity 84.2%; Pred. No. 2.8e-211;

Matches 567; Conservative 38; Mismatches 38; Indels 30; Gaps 5;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICSAREKLVWTVVYGVVPVWKEATTLFCASDAKAY 60

Db 1 MRATEIRKNYQHLWKGGTLLGLMLMICSAAEQRLWTVVYGVVPVWKEATTLFCASDAKAY 60

Qy 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 120

Db 61 DTEVHNWATHACVPTDPNPQEVKLENVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTN-LKNATNTKSSNWKEMDRGEIKNCSEFKV----- 159

Db 121 LTPLCVTLNCTDLRNATNTSSSWETMEKEIKNCSEFNITTSIRDKVQKEYALFYNLDDV 180

Qy 160 ----GAGKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDDKFKNGSGPCTNVST 215

Db 181 PIDNASYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDDKFKNGTGPCTNVST 240

Qy 216 VQCTHGIRPVVSTQLLNGSLAEEGVIRSENFDTNAKTIIVQLKESVEINCTRPNNT 275

Db 241 VQCTHGIRPVVSTQLLNGSLAEEEIVIRSENFDTNNAKTIIVQLNESVINCTRPNNT 300

Qy 276 KSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIVFKOSS 334

Db 301 K SINIGPGRALYTTGEIIGDIRQAHCNLSKTQWENTLEQIAIKLKEQFGNNKTIIFNPSS 360

Qy 335 GGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNTNGTITLPCR IKQIINRWQEVG 394

Db 361 GGDPEIVTHSFNCGGEFFYCNSTQLF--TWNDTRKLNNTGRNITLPCR IKQIINMWQEVG 418

Qy 395 KAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVK 454

Db 419 KAMYAPPPIRGQIRCSSNITGLLLTRDGGKD-TNGTEIFRPGGDMRDNRSELYKYKVVK 477

Qy 455 IEPLGVAPTAKARRVVQREKRAVTLGAMFLGFLGAAGSTMGAARSLTLTVQARQLLSGIVQ 514

Db 478 IEPLGVAPTAKARRVVQREKRAVGLGALFLGFLGAAGSTMGAASITLTVQARQLLSGIVQ 537

Qy 515 QONNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQLLGIWCSGKLICTTAVPW 573

Db 538 QONNLLRAIEAQHLLQLTVWGIKQLQARVLAVERYLRDQLLGIWCSGKLICTTTVPW 597

Qy 574 NASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLW 633

Db 598 NTSWSNKSLSNEIWDNMTWMEWEREIDNYTHIIYSLIEQSNQOEKNEQELLALDKWASLW 657

Qy 634 NWFDISKWLMYIK 646

Db 658 NWFDITKWLWYIK 670

RESULT 3

Tl2016

envelope glycoprotein - human immunodeficiency virus type 1 (strain scl4.3)

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C;Accession: Tl2016

R;McCutchan, F.E.; Sanders-Buell, E.; Salminen, M.O.; Carr, J.K.; Sheppard, W.H.

AIDS Res. Hum. Retroviruses 14, 329-337, 1998

A;Title: Diversity of the human immunodeficiency virus type 1 envelope glycoprotein in S

A;Reference number: Z17379; MUID:98178716; PMID:9519894

A;Accession: Tl2016

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-852 <MCC>

A;Cross-references: EMBL:U90934; NID:g2351783; PIDN:AAC59271.1; PID:g2351784

C;Genetics:

A;Gene: env

C;Superfamily: type E retrovirus env polyprotein

Query Match 87.1%; Score 3025.5; DB 2; Length 852;

Best Local Similarity 84.3%; Pred. No. 8.4e-211;

Matches 575; Conservative 34; Mismatches 34; Indels 39; Gaps 7;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICSAREKLVWTVVYGVVPVWKEATTLFCASDAKAY 60

Db 1 MRVKGIRKNQCQHLWKWGTMLLGLMLMICSAAEQRLVTVVYGVVPVWKEATTLFCASDAKAY 60

Qy 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 120

Db 61 DTEVHNWATHACVPTDPNPQEVVLKNVTENFNMMKNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTN-LKNATNTKSSNWKEMDRGEIKNCSEFKV----- 159

Db 121 LTPLCVTLNCTDYLNDTNTSSNGGMEGEIKNCSEFNITTRIGNKVQKEYALFYKLDV 180

Qy 160 ----GAGKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDDKFKNGSGPCTN 212

Db 181 VPIDNTTTSYRLINCNTSVITQACPKVSFEPIPIHYCTPAGFALLKCKDKKFKNGTGPCTN 240

Qy 213 VSTVQCTHGIRPVVSTQLLNGSLABEGVIRSENFDTNAKTIIVQLKESVEINCTRPN 272

Db 241 VSTVQCTHGIRPVVSTQLLNGSLABEEVIRSENFDTNAKTIIVQLNESVEINCTRPN 300

Qy 273 NTRKSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQ 332

```

Db      301 NTRKSIHIGPGQALYATGAIIGDIRQAHCNISRAKWNNTLKQIVKKLVQFGNKTIIFNQ 360
Qy      333 SSGDPEIVMHSFNCGGEFFYCNSQTOLFNSTW--NNTIGPNNT- ----NGTITLPCRIRKQ 385
Db      361 SSGDPEIVMHSFNCGGEFFYCNTTKLFNSTWMFNNTW--NDTEDTEGNGTITLPCRIRKQ 418
Qy      386 IINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDMRDNWRS 445
Db      419 IINRWQEVGKAMYAPPPIRGQIRCSSNITGILLTRDGGTN-NSTNETFRPGGGDMRDNWRS 477
Qy      446 ELYKYKVVKIEPLGVAPTAKARRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQA 505
Db      478 ELYKYKVVKIEPLGVAPTAKARRVVQREKRAVGIGALFLGFLGAAGSTMGAADVTLTVQA 537
Qy      506 RQLLSGIVQQONNLLRAIEAQHLLQLTVWGIK-LOARVLAVERYLKDQQLLGIWGCSGK 564
Db      538 RQLLSGIVQQONNLLRAIEAQHLLQLTVWGIKQLQARVLAVERYLRDQQLLGIWGCSGK 597
Qy      565 LICTTAVPNWNASWSKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELL 624
Db      598 LICTTTVPWNASWSKSLDKIWNNTWMEWEREIDNYTSLIYSLIEESQNQVQKNEQELL 657
Qy      625 ELDKWASLWNWFDISKWLWYIK 646
Db      658 ELDKWASLWSWFEDITKWLWYIK 679

RESULT 4
S13289
env protein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997
C;Accession: S13289
R;O'Brien, W.A.; Koyanagi, Y.; Namazie, A.; Zhao, J.Q.; Diagne, A.; Idler, K.; Zack, J.A
Nature 348, 69-73, 1990
A;Title: HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120
A;Reference number: S13288; MUID:91043044; PMID:2172833
A;Accession: S13289
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-847 <OBR>
C;Superfamily: type E retrovirus env polyprotein

Query Match      86.6%; Score 3008; DB 2; Length 847;
Best Local Similarity 84.6%; Pred. No. 1.5e-209;
Matches 572; Conservative 29; Mismatches 43; Indels 32; Gaps 6;

Qy      1 MRVKGIRKNYQHLWRGGTLLGLMLMICSAREKLVWTVVYGVVPVWKEATTTLFCASDAKAY 60
Db      1 MRVKGIRKSYQYLVKWKGTMLLGLMLMICSATEKLVWTVVYGVVPVWKEATTTLFCASDAKAY 60
Qy      61 DTEVHNVWATHACVPTDPNPQEI VLENTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db      61 DTEVHNVWATHACVPTDPNPQEVVLENTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Qy      121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db      121 LTPLCVTLNCKDV-NATNTTNDSEGTMERGEIKNCSFNITTSIRDEVQKEYALFYKLDVV 179
Qy      163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNCGPCTNV 213
Db      180 PIDNNNTSYRLISCDT SVITQACPKISFEPIPIHYCAPAGFAILKCNCKT FNGKGPCKNV 239
Qy      214 STVQCTHGIRPVVSTQLLNGSLAEEGWVIRSENFTDNAKTIIVQLKESVEINCTRPNNN 273
Db      240 STVDCTHGIRPVVSTQLLNGSLAEEVWVIRSDNFTNNAKTIIVQLKESVEINCTRPNNN 299
Qy      274 TRKSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNKTIIVFKQS 333
Db      300 TRKSIHIGPGRAFYTTGEEIIGDIRQAHCNISRAKWNNTLKQIVIKLREQFENKTIIVFNHS 359
Qy      334 SSGDPEIVMHSFNCGGEFFYCNSTQLFNSTW-NNTIGPNNTNG-TITLPCRIRKQIINRWQ 391
```

```

Db      360 SSGDPEIVMHSFNCGGEFFYCNSQTOLNNSTWNNNTSGSNNTTEGNTITLPCRIRKQFINMWQ 419
Qy      392 EVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDMRDNWRSELYKYK 451
Db      420 EVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGIN-ENGTEIFRPGGGDMRDNWRSELYKYK 478
Qy      452 VVKIEPLGVAPTAKARRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSG 511
Db      479 VVKIEPLGVAPTAKARRVVQREKRAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSG 538
Qy      512 IVQQONNLLRAIEAQHLLQLTVWGIK-LOARVLAVERYLKDQQLLGIWGCSGKLICTTA 570
Db      539 IVQQONNLLRAIEAQQRMLQLTVWGIKQLQARVLAVERYLGDQQLLGIWGCSGKLICTTA 598
Qy      571 VPWNASWSKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELLELDKWA 630
Db      599 VPWNASWSKSLDRIWNNTWMEWEREIDNYTSEIYTLIEESQNQOEKNEQELLELDKWA 658
Qy      631 SLWNWFDISKWLWYIK 646
Db      659 SLWNWFDITKWLWYIK 674

RESULT 5
VCLJAJ2
env polyprotein precursor - human immunodeficiency virus type 1 (isolate ARV-2)
N;Alternate names: coat polyprotein
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Accession: A03976
R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Shi
Science 227, 484-492, 1985
A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).
A;Reference number: A04003; MUID:85090453; PMID:2578227
A;Accession: A03976
A;Molecule type: DNA
A;Residues: 1-855 <SAN>
A;Cross-references: GB:K02007; NID:g328658; PIDN:AAB59882.1; PID:g328666
C;Genetics:
A;Gene: env
C;Superfamily: type E retrovirus env polyprotein
C;Keywords: AIDS; capsid protein; coat protein; glycoprotein; immunodeficiency; polyprot
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-509/Product: exterior membrane glycoprotein #status predicted <EXT>
F;510-855/Product: transmembrane glycoprotein #status predicted <TMM>
F;87,129,140,158,184,190,200,244,265,292,298,304,334,341,358,364,388,394,400,408,445,458,
F;610,624,636,815/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      85.9%; Score 2984; DB 1; Length 855;
Best Local Similarity 82.4%; Pred. No. 8.5e-208;
Matches 563; Conservative 34; Mismatches 48; Indels 38; Gaps 6;

Qy      1 MRVKGIRKNYQHLWRGGTLLGLMLMICSAREKLVWTVVYGVVPVWKEATTTLFCASDAKAY 60
Db      1 MKVKGTRRNYQHLWRWGTLLGLMLMICSATEKLVWTVVYGVVPVWKEATTTLFCASDARAY 60
Qy      61 DTEVHNVWATHACVPTDPNPQEI VLENTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db      61 DTEVHNVWATHACVPTDPNPQEVVLENTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Qy      121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db      121 LTPLCVTLNCTDLGKATNTSSNWKKEIKGEIKNCSFNITTSIRDKIQENALFRNLDDV 180
Qy      163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNCGSG 208
Db      181 PIDNASTTNTYRLIHCNRSVITQACPKVSFEPIPIHYCTPAGFAILKCNCKTFNGKG 240
Qy      209 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGWVIRSENFTDNAKTIIVQLKESVEINCT 268
Db      241 PCTNVSTVQCTHGIRPIVSTQLLNGSLAEEVWVIRSDNFTNNAKTIIVQLNESVAINCT 300
```


Db 118 PCVKLTPLCVSLKCTDLKNDTNTSSGRMIMEKEIKNCSFNISTSRGKVQKEYAFFY 177
QY 164 -----LINCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNKKFNGSG 208
Db 178 KLDIIPIDNDTTSYTLTSCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNKTFNGTG 237
QY 209 PCTNVSTVQCTHGIRPVSTQLLNGSLAEEGVIRSENFTDNAKTIIVQLKESVEINCT 268
Db 238 PCTNVSTVQCTHGIRPVSTQLLNGSLAEDEVIRSANFTDNAKTIIVQLNQSVINCT 297
QY 269 RPNNTRKSTI--GPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFG-N 325
Db 298 RPNNTRKSTRIQRPGRFVTIGK-IGNRQAHCNISRAKWNNTLKQIDSKLREQFGNN 356
QY 326 KTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNN-----TIGPNNTNG--TITL 379
Db 357 KTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFSNTWSTKGSNNTGSDTITL 416
QY 380 PCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDM 439
Db 417 PCRIKQIINMWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNS--NNESEIFRPGGDM 475
QY 440 RDNWRSELYKVKVVKIEPLGVAPTAKKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSL 499
Db 476 RDNWRSELYKVKVVKIEPLGVAPTAKKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASM 535
QY 500 TLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGK-LQARVLAVERYLKDQQLLGI 558
Db 536 TLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGKIQLOARILAVERYLKDQQLLGI 595
QY 559 WCGSGKLICTTAVPWNASWSNKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQOEK 618
Db 596 WCGSGKLICTTAVPWNASWSNKSLEQIWNNTWMEWDREINNTYLSLHSLIEESQNQOEK 655
QY 619 NEQELLELDKWASLWNWFDISKWLWYIK 646
Db 656 NEQELLELDKWASLWNWFNITNWLWYIK 683

RESULT 8
VCLJYBR
env polyprotein - human immunodeficiency virus type 1 (isolate BR)
N:Alternate names: coat polyprotein
N:Contains: coat protein gp120; coat protein gp41
C:Species: human immunodeficiency virus type 1, HIV-1
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Feb-1997
C:Accession: A31667
R:Anand, R.; Thayer, R.; Srinivasan, A.; Nayyar, S.; Gardner, M.; Luciw, P.; Dandekar, S
Virology 168, 79-89, 1989
A:Title: Biological and molecular characterization of human immunodeficiency virus (HIV-
A:Reference number: A94389; MUID:89085613; PMID:2789516
A:Accession: A31667
A:Molecule type: DNA
A:Residues: 1-852 <ANA>
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: capsid protein; coat protein; polyprotein; transmembrane protein
F:1-516/Product: coat protein gp120 #status predicted <CP1>
F:517-852/Product: coat protein gp41 #status predicted <CP2>

Query Match 84.6%; Score 2938.5; DB 1; Length 852;
Best Local Similarity 81.6%; Pred. No. 1.7e-204;
Matches 556; Conservative 41; Mismatches 47; Indels 37; Gaps 8;

QY 1 MRVKGIRKNYQHLWR-GGTLLLGLMLMICSAREKLVWTVVYGVVWKEATTTLFCASDAKA 59
Db 1 MRVKGIRKNYQHLWRGGMMLLGILMICSATDKLVWTVVYGVVWKEANTTLFCASDAKA 60
QY 60 YDTEVHNWATHACVPTDPNPQEIIVLENVTENFNMWKNNMVEQMHEDIISLWDQSLKPCV 119
Db 61 YDTEIHNWATHACVPTDPNPQELVMGNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCV 120

QY 120 KLTPLCVTLHCTNLKNAATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db 121 KLTPLCLTLNCHDF-NATNATNSGKMMEGEMKNCNFNITTSIRDKMQKEYALFYKLDI 179
QY 163 -----KLINCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNKKFNGSGPC 210
Db 180 VPIDNDKTNTRYRLISCNSTSVITQACPKVTFEPIPIHYCAPAGFAILKCNKKFNGTGPC 239
QY 211 TNVSTVQCTHGIRPVSTQLLNGSLAEEGVIRSENFTDNAKTIIVQLKESVEINCTRP 270
Db 240 TNVSTVQCTHGIRPVSTQLLNGSLAEDEVIRSENFTNNVKTIIIVQLNESVEINCTRP 299
QY 271 NNNTRKSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVE 330
Db 300 NNNTRKRITMGPRVYTTGQIIGDIRRAHCNLSRSKWENTLKQIVTKLRVQFKNKTIVE 359
QY 331 KQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTW-NNTIGPNNTNGT--ITLPCRKQII 387
Db 360 NRSSGGDPEIVMHSFNCGGEFFCNTTQLFNSTWYRNTTG-NITEGNSPITLPCRKQII 418
QY 388 NRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNT-TEIFRPGGDMRDNWRSE 446
Db 419 NMWQEVGKAMYAPPPIRGQIKCSSNITGLLLTRDGGNNNETTDEIFRPGGDMRDNWRSE 478
QY 447 LYKYKVVVKIEPLGVAPTAKKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQAR 506
Db 479 LYKYKVVVKIEPLGVAPTAKKRRVVQREKRAVGLGALFLGFLGAAGSTMGAASLTLTVQAR 538
QY 507 QLLSGIVQQNNLLRAIEAQOHLQLTVWGK-LQARVLAVERYLKDQQLLGWCGSGKL 565
Db 539 LLLSGIVQQNNLLMAIEAQOHLMLTLTVWGKIQLOARVLAVERYLKDQQLLGWCGSGKL 598
QY 566 ICTTAVPWNASWSNKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELLE 625
Db 599 ICTTAVPWNASWSNKSLSLDIWNNTWMEWEREIDNYTNLIYSLIEDSQIQOEKNEKELLE 658
QY 626 LDKWASLWNWFDISKWLWYIK 646
Db 659 LDKWASLWNWFNITNWLWYIK 679

RESULT 9
VCLJH4
env polyprotein - human immunodeficiency virus type 1 (isolate CDC-451)
N:Alternate names: coat polyprotein
N:Contains: coat protein gp120; coat protein gp41
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C:Accession: C25523
R:Desai, S.M.; Kalyanaraman, V.S.; Casey, J.M.; Srinivasan, A.; Andersen, P.R.; Devare, S
Proc. Natl. Acad. Sci. U.S.A. 83, 8380-8384, 1986
A:Title: Molecular cloning and primary nucleotide sequence analysis of a distinct human
A:Reference number: A94136; MUID:87041461; PMID:3490666
A:Accession: C25523
A:Molecule type: DNA
A:Residues: 1-868 <DES>
A:Cross-references: GB:M13137; NID:g326460; PIDN:AAA44311.1; PID:g326467
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: capsid protein; coat protein; glycoprotein; polyprotein; transmembrane protei
F:1-521/Product: coat protein gp120 #status predicted <GP1>
F:522-868/Product: coat protein gp41 #status predicted <GP2>
F:89,138,139,142,166,195,198,208,245,252,273,287,300,312,342,349,365,371,395,405,409,459,

Query Match 84.6%; Score 2937.5; DB 1; Length 868;
Best Local Similarity 79.5%; Pred. No. 2e-204;
Matches 556; Conservative 37; Mismatches 47; Indels 59; Gaps 8;

QY 1 MRVKGIRKNYQHLWRGGTLLGLMLMICSAREKLVWTVVYGVVWKEATTTLFCASDAKAY 60
Db 3 MRAKGIRKNQCHLWRWGTMLLGLMLMICSAAANLVWTVVYGVVWKEATTTLFCASDAKAY 62

Qy	61	DTEVHNWATHACVPTDPNPQEI	VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK	120	
Db	63	DTEAHNVWATHACVPTPNPQEV	VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK	122	
Qy	121	LTPLCVTLHCT--NLKNATNTKSSN	-----WKEMDRGEIKNCSFKV	GAG----- 162	
Db	123	LTPLCVTLNCTDLNTNTNTTNT	TELSIIVWEQRGKGMERNCSFNITTSIRDKVQREYALF	182	
Qy	163	-----	KLINCN	TSVITQACPKVSFEPIPIHYCAPAGFAILKCNDK 202	
Db	183	YKLDVEPIDDKNTTNTTKYRL	INCNTSVITQACPKVSFEPIPIHYCTPTGFALLKCNDK	242	
Qy	203	KFNGSGPCTNVSTVQCTHGIRPV	VSTQLLNGSLAEEGVVIRSENFTDNAKTIIVQLKES	262	
Db	243	KFNGTGPCTNVSTVQCTHGIRPV	VSTQLLNGSLAEEVVIRSENFTNNAKTIIVQLNVS	302	
Qy	263	VEINCTRPNNNTRKSITIGP	GRAFYATGDIIGDIRQAH	CNISGEKWNNTLKQIVTKLQAO 322	
Db	303	VEINCTRPNNHTRKRVTLGP	RVWYTTGEILGNIRQAH	CNISRAQWNTLQQA	IATTLREQ 362
Qy	323	FGNKTIIVFKQSSGGDPEI	VMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNTNGT	----- 376	
Db	363	FGNKTIAPNQSSGGDPEI	VMHSFNCGGEFFYCNSTQLFNSAMNVT	-----SNGTWSVTRK 417	
Qy	377	-----ITLPCRIKQI	INRWQEVGKAMYAPP	IRGQIRCSSNITGLLLTRDGGKEISNTT 429	
Db	418	QKDTGDIITLPCRIKQI	INRWQVVGKAMYALPIKGLIRCSSNITGLLLTRDGGGE-NQTT	476	
Qy	430	EIFRPGGDMRDNRSELYKY	VVKVVKIEPLGVAPTAKRRVVQREKRAV-TLGAMFLGFLG	488	
Db	477	EIFRPGGDMRDNRSELYKY	VVKVVKIEPLGVAPTAKRRVVQREKRAVGMFLGFLG	536	
Qy	489	AAGSTMGARSLTLTVQARQLLSGIV	QQNNLLRAIEAQOHLQLTVWG	IK-LQARVLA 547	
Db	537	AAGSTMGATSMALTVQARQLLSGIV	QQNNLLRAIKAQOHLQLTVWG	IKQLQARILA 596	
Qy	548	RYLKDQQLGIWCGSGKLICTTAV	PWNASWSNKS	LQIWNMTWMEWEREIDNYTNLIYT 607	
Db	597	RYLKDQQLLGFWCGSGKLICTTAV	PWNASWSNKTLDQIWNMTWMEWDREIDNYTHLIYT	656	
Qy	608	LIEESNQOQEKNEQELLELDKWASL	WNWFDISKWLWYIK	646	
Db	657	LIEESNQOQEKNOQELLQDKWASL	WTSWDSITKWLYIK	695	

RESULT 10
VCLJLV
env polyprotein precursor - human immunodeficiency virus type 1 (isolate LAV-1a)
N:Alternate names: coat polyprotein
C:Species: human immunodeficiency virus type 1, HIV-1
A:Note: host Homo sapiens (man)
C>Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 16-Jul-1999
C:Accession: A03975
R:Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A:Title: Nucleotide sequence of the AIDS virus, LAV.
A:Reference number: A90866; MUID:85099333; PMID:2981635
A:Accession: A03975
A:Molecule type: DNA
A:Residues: 1-861 <WAI>
A:Cross-references: GB:K02013; NID:g326417; PIDN:AAB59751.1; PID:g326424
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: AIDS; capsid protein; coat protein; glycoprotein; immunodeficiency; polypro
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-516/Product: exterior membrane glycoprotein #status predicted <EXT>
F:517-861/Product: transmembrane glycoprotein #status predicted <TM>
F:88,136,141,146,161,165,191,202,239,246,267,281,294,300,306,337,344,361,391,397,402,411
F:616,621,630,642,679,755,821/Binding site: carbohydrate (Asn) (covalent) #status predic
Query Match 84.4%; Score 2933; DB 1; Length 861;
Best Local Similarity 81.8%; Pred. No. 4.2e-204;

Matches	567;	Conservative	33;	Mismatches	41;	Indels	52;	Gaps	12;
Qy	1	MRVKGIRKNYQHLWR----	GGTLLGLMLMICS	AVEKLVWTVVYGV	PVWKEATTTLFC	ASD	56		
Db	1	MRVK--EKYQHLWRWG	WKGTMLLGILMICS	ATEKLVWTVVYGV	PVWKEATTTLFC	ASD	57		
Qy	57	AKAYDTEVHNWATHACVPTDP	NPQEI	VLENVTENFNMWKNM	VEQM	HEDIISLWDQSLK	116		
Db	58	AKAYDTEVHNWATHACVPTDP	NPQEV	VLNVNVTENFNMWKN	DMVEQM	HEDIISLWDQSLK	117		
Qy	117	PCVKLTPLCVTLHCTNLK	NATNTKSSNWKE-----	MDRGEIKNCSFK	VGA---GK----	163			
Db	118	PCVKLTPLCVSLKCTDL	GNATNTSSNTSSSGEM	MMERGEIKNCSFN	ISTSIRGKVQKE	177			
Qy	164	-----	LINCNTSVITQACPK	VSFEPIPIHYCAPAG	FAILKCNDKK	203			
Db	178	YAFFYKLDIIPIDND	TTSYTLTSCNTSVITQ	ACPKVSFEPIPIHYC	APAGFAILKCNNKT	237			
Qy	204	FNGSGPCTNVSTVQCTH	GIRPVVSTQLLNGSLA	EEGVVRSENFTD	NAKTIIVQLKESV	263			
Db	238	FNGTGPCTNVSTVQCTH	GIRPVVSTQLLNGSLA	EEVVRSANFTD	NAKTIIVQLNQSV	297			
Qy	264	EINCTRPNNNTRKSITI--	GPGRAFYATGDIIG	DIRQAH	CNISGEKWNNTLKQIVTKLQ	QA 321			
Db	298	EINCTRPNNNTRKSIRI	QRGPGRAFTVIGK-IGN	MRQAH	CNISRAKWNATLKQI	ASKLRE 356			
Qy	322	QFG-NKTIIVFKQSSGG	DPEI	VMHSFNCGGEFFYCN	STQLFNSTWN	N---TIGPNNTNG- 375			
Db	357	QFGNNKTIIFKQSSGG	DPEI	VTHSFNCGGEFFYCN	STQLFNSTW	STEGSNNTGS 416			
Qy	376	-TITLPCRIKQI	INRWQEVGKAMYAPP	IRGQIRCSSNITGLLL	TRDGGKEISNTTEIFRP	434			
Db	417	DTITLPCRIKQFIN	MWQEVGKAMYAPPIS	GQIRCSSNITGLLL	TRDGGNN-NNGSEIFRP	475			
Qy	435	GGGMRDNWRSELYKY	VVKVVKIEPLGVAPTAK	RRRVVQREKRAV	TLGAMFLGFLGAAGSTM	494			
Db	476	GGGMRDNWRSELYKY	VVKVVKIEPLGVAPTAK	RRRVVQREKRAV	GIGALFLGFLGAAGSTM	535			
Qy	495	GARSLTLTVQARQLLS	GIVQQNNLLRAIEAQ	QHLLQLTVWG	IK-LQARVLA	VERYLKDQ 553			
Db	536	GARSMTLTVQARQLLS	GIVQQNNLLRAIEAQ	QHLLQLTVWG	IKQLQARILA	VERYLKDQ 595			
Qy	554	QLLGIWCGSGKLICTT	AVPWNASWSNKS	LDQIWNMTWME	WEREIDNYTNLIYTLIEESQ	613			
Db	596	QLLGIWCGSGKLICTT	AVPWNASWSNKS	LEQIWNMTWME	WDREINNYTSLI	HSLIEESQ 655			
Qy	614	NQOEKNEQELLELD	KWASLWNWFDISK	WLYIK	646				
Db	656	NQOEKNEQELLELD	KWASLWNWENITN	WLYIK	688				
RESULT 11									
S33985									
env polyprotein - human immunodeficiency virus type 1									
C:Species: human immunodeficiency virus type 1, HIV-1									
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999									
C:Accession: S33985									
R:Carlini, F.									
submitted to the EMBL Data Library, November 1991									
A:Reference number: S33979									
A:Accession: S33985									
A:Status: preliminary									
A:Molecule type: mRNA									
A:Residues: 1-851 <CAR>									
A:Cross-references: EMBL:Z11530; NID:g60192; PIDN:CAA77628.1; PID:g60199									
C:Superfamily: type E retrovirus env polyprotein									
Query Match									
Best Local Similarity									
84.3%; Score 2928; DB 2; Length 851;									
82.5%; Pred. No. 9.5e-204;									
Matches 564; Conservative 34; Mismatches 42; Indels 44; Gaps 11;									
Qy	1	MRVKGIRKNYQHL----	WRG	GTLLGLMLMICS	AVEKLVWTVVYGV	PVWKEATTTLFC	ASD	56	

Db	1	MRVK---	EKYQHLWRGW	RWGTMLL	GMLMICS	ATEKLWTVVYGV	PVWKEAT	TTLFCASD	57					
Qy	57	AKAYD	TEVHNW	WATHAC	VPDNP	QEI	VLEN	TENFNM	KNNMVE	QMHE	DIISLW	DQSLK	116	
Db	58	AKAYD	TEVHNW	WATHAC	VPDNP	QEVV	LVN	TENFNM	KNDMVE	QMHE	DIISLW	DQSLK	117	
Qy	117	PCVKLT	PCLCVTL	HCTNL	KNATNT	KSSNWKE	-MDRGEI	KNC	SFKVGA	-----	163			
Db	118	PCVKLT	PCLCVSLK	CTDLK	NDTNTN	SSGRM	IMEKEI	KNC	SFNIST	SKRGK	VQKEY	AFY	177	
Qy	164	-----	LINCNT	SVITQ	ACPKV	SFEPI	PIHYC	APAG	FAILK	CNDK	KKFNG	SG	208	
Db	178	KLDI	IPIDND	TTSY	TLTSC	NTSVIT	QACPKV	SFEPI	PIHYC	APAG	FAILK	CNNKT	FN	237
Qy	209	PCTNV	STVQ	CTHG	IRPV	VSTQ	LLNG	SLAE	EGV	IRSEN	FTD	NAKTI	I	268
Db	238	PCTNV	STVQ	CTHG	IRPV	VSTQ	LLNG	SLAE	EEV	IRSEN	FTD	NAKTI	I	297
Qy	269	RPNN	TRKS	ITI	--GP	GRA	FATG	DI	IGD	IRQA	HCNIS	G	325	
Db	298	RPNN	TRKKIR	I	Q	R	G	P	F	A	T	G	356	
Qy	326	KTIV	FKQ	SSG	DP	PEI	VMH	S	F	NC	G	E	383	
Db	357	KTII	IFK	Q	SSG	DP	PEI	V	H	S	F	NC	415	
Qy	384	KQI	INR	WQ	EV	G	K	A	M	A	P	I	443	
Db	416	KQI	INM	WQ	EV	G	K	A	M	A	P	I	474	
Qy	444	RSE	LYK	V	K	V	K	I	E	P	L	G	503	
Db	475	RSE	LYK	V	K	V	K	I	E	P	L	G	534	
Qy	504	QAR	QL	SGI	V	Q	Q	N	L	L	R	A	562	
Db	535	QAR	QL	SGI	V	Q	Q	N	L	L	R	A	594	
Qy	563	GK	L	I	C	T	T	A	V	P	N	A	622	
Db	595	GK	P	I	C	T	T	A	V	P	N	A	654	
Qy	623	L	L	E	L	D	K	W	A	S	L	N	646	
Db	655	L	L	E	L	D	K	W	A	S	L	N	678	
RESULT 12														
VCLJVL														
env polyprotein precursor - human immunodeficiency virus type 1 (isolate LV)														
N;Alternate names: coat polyprotein														
C;Species: human immunodeficiency virus type 1, HIV-1														
A;Note: host Homo sapiens (man)														
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 16-Jul-1999														
C;Accession: A03974														
R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.														
Nature 313, 450-458, 1985														
A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi														
A;Reference number: A93355; MUID:85111157; PMID:2982104														
A;Accession: A03974														
A;Molecule type: DNA														
A;Residues: 1-856 <MUE>														
A;Cross-references: GB:K02083; NID:g555008; PIDN:AAB59873.1; PID:g328559														
C;Genetics:														
A;Gene: env														
C;Superfamily: type E retrovirus env polyprotein														
C;Keywords: AIDS; capsid protein; coat protein; glycoprotein; immunodeficiency; polypro														
F;1-30/Domain: signal sequence #status predicted <SIG>														
F;31-511/Product: exterior membrane glycoprotein #status predicted <EXT>														
F;512-856/Product: transmembrane glycoprotein #status predicted <TMM>														
F;88,136,141,156,160,186,197,230,234,241,262,276,289,295,301,332,339,356,386,392,397,406														
F;611,616,625,637,674,750,816/Binding site: carbohydrate (Asn) (covalent) #status predic														

Query Match		84.2%; Score 2926.5; DB 1; Length 856;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Best Local Similarity		82.1%; Pred. No. 1.2e-203;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Matches 565; Conservative 35; Mismatches 41; Indels 47; Gaps 12;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
QY	1	MRVKGIRK	NYQHL-----WRGGTLL	LGMLMICS	AVEK	LWTVVY	GV	VPVW	KEATTT	LF	CASD	56																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
DB	1	MRVK---	EKYQHL	WRGWRG	WTML	LGMLMICS	ATEK	LWTVVY	GV	VPVW	KEATTT	LF	CASD	57																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
QY	57	AKAYDTE	VHNV	WATHAC	VP	TDPNP	QEI	VLEN	VTFN	FMKNN	MVEQ	MHEDI	IS	LWDQ	SLK	116																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
DB	58	AKAYDTE	VHNV	WATHAC	VP	TDPNP	QEV	L	VNV	TENF	NMKN	D	MVEQ	MHEDI	IS	LWDQ	SLK	117																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
QY	117	PCVKLT	PL	CV	T	LH	C	T	N	L	K	N	A	T	N	K	S	N	W	K	E	-	M	D	R	G	E	I	K	N	C	S	F	K	V	G	A	---	G	K	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---</

Qy 1 MRVKGIRKNYQHLWR-----GGTLLGLMLMICS AVEKLVTVVYGVVPVWKEATTTILFCASD 56
Db 1 MRVK-----EKYQHLRWGWKGWMLLGLMILMICSATEKLVTVVYGVVPVWKEATTTILFCASD 57
Qy 57 AKAYDTEVHNWATHACVPTDPNPQEIIVLENTNFNMKNMVEQMHEDIISLWDQSLK 116
Db 58 AKAYDTEVHNWATHACVPTDPNPQEVVLVNVTFENFMKNMVEQMHEDIISLWDQSLK 117
Qy 117 PCVKLTPLCVTLHCTNLKNATNTKSNWKE-MDRGEIKNCSEKVGAG----- 162
Db 118 PCVKLTPLCVSLKCTDLKNDTNTSSGRMIMEKEGEIKNCSEFNISTSIRDKVQKEYAFY 177
Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPC 210
Db 178 KLDIVPIDNTSYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNKTFNGTGPC 237
Qy 211 TNVSTVQCTHGIRPVVSTQLLNGSLAEGVVIRSENFNTDNAKTIIVQLKESVEINCTRP 270
Db 238 TNVSTVDCTHGIRPVVSTQLLNGSLAEEDVVIRSANFTDNAKTIIVQLNTSVEINCTRP 297
Qy 271 NNTRKSTI--GPGRAFATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAFG-NKT 327
Db 298 NNTRKSTIRIQRPGRAFVTIG-IGNMRQAHCNISRAKWNATLKQIASKLREQFGNNKT 356
Qy 328 IVFKQSSGGDPEIVMHSFNCGGEFFVCNSTQLFNSTWNN----TIGPNNTNG--TITLPC 381
Db 357 IIFKQSSGGDPEIVTHSFNCGGEFFVCNSTQLFNSTWFNSTWTEGSNNTGSDTITLPC 416
Qy 382 RIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDMRD 441
Db 417 RIKQFINRWQEVGKAMYAPPISGQIRCSSNITGLLLTRDGGNN--NGSEIFRPGGGDMRD 475
Qy 442 NWRSELYKYKVKVIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAAGSTMGARSLTL 501
Db 476 NWRSELYKYKVKVIEPLGVAPTAKRRVVQREKRAVGIGALFLGLGAAGSTMGCTSMTL 535
Qy 502 TVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGIK-LQARVLAVERYLKDQQLLGIWG 560
Db 536 TVQARQLLSDIVQQNNLLRAIEAQOHLQLTVWGIKQLQARILAVERYLKDQQLLGIWG 595
Qy 561 CSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQKEKNE 620
Db 596 CSGKLICTTAVPWNASWSNKSLEQIWNMTWMEWDREINNYTSLIHSLIEESQNQKEKNE 655
Qy 621 QELLELDKASLWNWFDISKWLWYIK 646
Db 656 QELLELDKASLWNWFNITNLWLWYIK 681

RESULT 14
VCLJKB
env polyprotein precursor - human immunodeficiency virus type 1 (strain KB-1-gp32)
N:Alternate names: coat polyprotein
N:Contains: coat protein gp120; coat protein gp32
C:Species: human immunodeficiency virus type 1, HIV-1
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 01-Mar-1996
C:Accession: B42995
R:Shimizu, H.; Hasebe, F.; Tsuchie, H.; Morikawa, S.; Ushijima, H.; Kitamura, T.
Virology 189, 534-546, 1992
A:Title: Analysis of a human immunodeficiency virus type 1 isolate carrying a truncated
A:Reference number: A42995; MUID:92351552; PMID:1322587
A:Accession: B42995
A:Molecule type: mRNA
A:Residues: 1-729 <SH1>
A:Cross-references: GB:S41266; GB:D01206
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F:1-689/Domain: extracellular #status predicted <EXT>
F:1-33/Domain: signal sequence #status predicted <SIG>
F:17-33/Region: hydrophobic #status predicted

F:34-517/Product: coat protein gp120 #status predicted <CP1>
F:514-517/Region: cleavage processing #status predicted
F:518-729/Product: coat protein gp32 #status predicted <CP2>
F:518-534/Region: hydrophobic #status predicted
F:690-711/Domain: transmembrane #status predicted <TM1>
F:712-729/Domain: intracellular #status predicted <INT>
F:93,141,145,146,163,191,192,237,241,248,269,283,308,338,345,361,367,397,403,408,414,
Query Match 83.7%; Score 2908.5; DB 1; Length 729;
Best Local Similarity 79.0%; Pred. No. 2e-202;
Matches 549; Conservative 42; Mismatches 49; Indels 55; Gaps 8;
Qy 1 MRVKGIRKNYQHLW-----RGGTLLGLMLMICS AVEKLVTVVYGVVPVWKEATTTILFCA 54
Db 1 MRVKEIRKNYQHLWRWGMILRWGTMLLGLMLMICSAAEQLVTVVYGVVPVWKEATTTILFCA 60
Qy 55 SDAKAYDTEVHNWATHACVPTDPNPQEIIVLENTNFNMKNMVEQMHEDIISLWDQS 114
Db 61 SDAKAYDTEAHNVWATHACVPTDPNPQEVVLVNVTFENFMKNMVEQMHENIISLWDQS 120
Qy 115 LKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEKVGAG----- 162
Db 121 LKPCVKLTPLCVTLHCTDLRNTTNNSSIEEKM-KGEIKNCSEFNVTNIRDKVQKEYALF 179
Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKK 203
Db 180 YKLDVVPIDNNDNSTNTCYRLISCNTSVITQACPKVSFEPIPIHYCTPAGFALLKCNKKT 239
Qy 204 FNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEGVVIRSENFNTDNAKTIIVQLKESV 263
Db 240 FNGTGPCKNVSTVQCTHGIRPVVSTQLLNGSLAEGVVIRSENFNTDVKTIIVQLNETV 299
Qy 264 EINCTRPNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAF 323
Db 300 KINCIRPNNKTRKRVTMGPGRVYTTGEIIGDIRQAHCNISRAEWNKLTLEQIANLKRQF 359
Qy 324 GNKTIVFKQSSGGDPEIVMHSFNCGGEFFVCNSTQLFNST-----WNNTIGPNNTNGT 376
Db 360 ENKTIVFNQSSGGDPEIVMHSFNCGGEFFVCDSQLFNSTHLSNGTWNNGT-GPEN---- 414
Qy 377 ITLPCRQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTT----EIF 432
Db 415 ITLPCRQIIVNMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGTQNNNTSSIEIF 474
Qy 433 RPGSGDMRDNRSELYKYKVKVIEPLGVAPTAKRRVVQREKRAVTLGAMFLGLGAAGS 492
Db 475 RPGSGDMRDNRSELYKYKVKVIEPLGVAPTAKRRVVQREKRAVGIGAVFLGLGAAGS 534
Qy 493 TMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGIK-LQARVLAVERYLK 551
Db 535 TMGAAAVTLTVQARQLLPQIVQQNNLLRAIDAOQHLLQLTVWGIKQLQARVLAVERYLK 594
Qy 552 DQQLGIWCGSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEE 611
Db 595 DQQLGIWCGSGKFICTTAVPWNSTWSNKSFEIWDNMTWMEWEREINNYTNLIYNLIEE 654
Qy 612 SQNOQEKNEQELLELDKASLWNWFDISKWLWYIK 646
Db 655 SQNOQEKNEQDLLALDKWDSLWNWFSITKWLWYIK 689

RESULT 15
VCLJKB
env polyprotein precursor - human immunodeficiency virus type 1 (strain KB-1-gp41)
N:Alternate names: coat polyprotein
N:Contains: coat protein gp120; coat protein gp41
C:Species: human immunodeficiency virus type 1, HIV-1
A:Note: host Homo sapiens (man)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 01-Mar-1996
C:Accession: A42995
R:Shimizu, H.; Hasebe, F.; Tsuchie, H.; Morikawa, S.; Ushijima, H.; Kitamura, T.
Virology 189, 534-546, 1992
A:Title: Analysis of a human immunodeficiency virus type 1 isolate carrying a truncated t

A:Reference number: A42995; MUID:92351552; PMID:1322587
A:Accession: A42995
A:Molecule type: mRNA
A:Residues: 1-861 <SHI>
A:Cross-references: GB:S41266; GB:D01206
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F:1-689/Domain: extracellular #status predicted <EXT>
F:1-33/Domain: signal sequence #status predicted <SIG>
F:17-33/Region: hydrophobic #status predicted
F:34-517/Product: coat protein gp120 #status predicted <CP1>
F:514-517/Region: cleavage processing #status predicted
F:518-861/Product: coat protein gp41 #status predicted <CP2>
F:518-534/Region: hydrophobic #status predicted
F:690-711/Domain: transmembrane #status predicted <TM1>
F:712-861/Domain: intracellular #status predicted <INT>
F:756-772/Region: hydrophobic #status predicted
F:93,141,145,146,163,191,192,237,241,248,269,283,296,308,338,345,361,367,397,403,408,414

Search completed: February 25, 2004, 14:19:20
Job time : 20.3842 secs

Query Match		83.7%;	Score 2908.5;	DB 1;	Length 861;
Best Local Similarity		79.0%;	Pred. No. 2.5e-202;		
Matches 549;		Conservative 42;	Mismatches 49;	Indels 55;	Gaps 8;
QY	1	MRVKGIRKNYQHLW-----RGGTLLGLMLICSAVEKLVWTVVYGVVPWK	EA	TTTLFCA	54
Db	1	MRVKEIRKNYQHLRWGIMLRWGTMLLGLMLICSAAEQLWTVVYGVVPWK	EA	TTTLFCA	60
QY	55	SDAKAYDTEVHNWATHACVPTDPNPQEI	VLENVTENFNMKNMNV	EQMHEDIISLWDQS	114
Db	61	SDAKAYDTEAHNVWATHACVPTDPNPQEV	VLNVVTENFNMKNMNV	EQMHENIISLWDQS	120
QY	115	LKPCVKLTPLCVTLHCTNLKNATNTKSNW	KEMDRGEIKNC	SFKVGAG-----	162
Db	121	LKPCVKLTPLCVTLHCTDLRNTNNSSIE	EKM-KGEIKNC	SFNVTTNIRD	KVQKEYALF 179
QY	163	-----KLINCNTSVITQACPKVS	FEPPIHYCAPAG	FAILKCNDKK	203
Db	180	YKLDVVPIDNNDNSTNTCYRLISCDT	SVITQACPKVS	FEPPIHYCTPAG	FALLKCNNKT 239
QY	204	ENGSGPCTNVSTVQCTHGIRPVVSTQL	LNGSLAEEGV	IRSENFTDNAKTIIVOLKESV	263
Db	240	FNGTGPCKNVSTVQCTHGIRPVVSTQL	LNGSLAEEGV	IRSENFTDNVKTIIIVQLNETV	299
QY	264	EINCTRPNNNTRKSITIGPGRAFYATG	DIIDIROAH	CNISGEKWNNTLKQIVTKLQAF	323
Db	300	KINCIRPNNKTRKRVTMGPGRVYTTG	EIIIDIROAH	CNISRAEWNKTLEQIANKLKQF	359
QY	324	GKNTIVFKQSSGGDPEIVMHSFNC	GGEFFYCNSTQLF	NST-----WNNTIGPNNTNGT	376
Db	360	ENKTIIVFNQSSGGDPEIVMHNFC	GGEFFYCDSSQLF	NSTHLSNGTWMNGT-GPEN----	414
QY	377	ITLPCR	IKQIINRWQEVGKAMYAPP	IRGQIRCSSNITGLLLTRDGGKEISNTT----	EIF 432
Db	415	ITLPCR	IKQIVNMWQEVGKAMYAPP	IRGQIRCSSNITGLLLTRDGGNTQNNNTNSSIEIF	474
QY	433	RPGGDMRDNRSELYKYKVVKIEPL	GVAPTAKRRVVQREK	RAVTLGAMFLGFLGAAGS	492
Db	475	RPGGDMRDNRSELYKYKVVKIEPL	GVAPTAKRRVVQREK	RAVGIGAVFLGFLGAAGS	534
QY	493	TMGARSLTLTVQARQLLSGIVQQ	NNLLRAIEAQOHL	LQLTVWGIK-LQARVLAVERYLK	551
Db	535	TMGAAAVTLTVQARQLLP	GIVQQNNLLRAIDAQOHL	LQLTVWGIKQLQARVLAVERYLK	594
QY	552	DQQLLG	INGCSGKLICTTAVPWNASW	SNKSLDQIWNNTMTMEWEREIDNYTNLIYTLIEE	611
Db	595	DQQLMG	INGCSGKFICTTAVPWNTS	WNKSFNEIWDNMTMEWEREINNYTNLIYNLIEE	654
QY	612	SONQOEKNEQELLELDKWASL	WNWFDISKWLWYIK	646	
Db	655	SONQOEKNEQDLLALDKWDSL	WNWFSITKWLWYIK	689	

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:03:35 ; Search time 11.2348 Seconds
(without alignments)
2994.034 Million cell updates/sec

Title: US-09-891-609A-2
Perfect score: 3474
Sequence: 1 MRVKGIRKNYQHLWRGGTLL.....DKWASLWNWFDISKWLWYIK 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3422	98.5	847	1 ENV_HV1S1	P19550 human immun
2	3032	87.3	843	1 ENV_HV1Y2	P35961 human immun
3	3002.5	86.4	848	1 ENV_HV1JR	P20871 human immun
4	2997.5	86.3	856	1 ENV_HV1SC	P05878 human immun
5	2984	85.9	855	1 ENV_HV1A2	P03378 human immun
6	2947	84.8	865	1 ENV_HV1RH	P04579 human immun
7	2941.5	84.7	852	1 ENV_HV1BN	P12488 human immun
8	2939.5	84.6	856	1 ENV_HV1B1	P03375 human immun
9	2937.5	84.6	868	1 ENV_HV1C4	P05879 human immun
10	2935	84.5	867	1 ENV_HV1J3	P12489 human immun
11	2933	84.4	861	1 ENV_HV1BR	P03377 human immun
12	2929	84.3	855	1 ENV_HV1OY	P20888 human immun
13	2926.5	84.2	856	1 ENV_HV1PV	P03376 human immun
14	2922.5	84.1	856	1 ENV_HV1H2	P04578 human immun
15	2918	84.0	851	1 ENV_HV1B8	P04582 human immun
16	2913.5	83.9	852	1 ENV_HV1S3	P19549 human immun
17	2908.5	83.7	861	1 ENV_HV1KB	P31819 human immun
18	2904	83.6	856	1 ENV_HV1MN	P05877 human immun
19	2896.5	83.4	856	1 ENV_HV1H3	P04624 human immun
20	2894.5	83.3	856	1 ENV_HV1W1	P31872 human immun
21	2891.5	83.2	856	1 ENV_HV1LW	Q70626 human immun
22	2890	83.2	847	1 ENV_HV1W2	P05880 human immun
23	2885.5	83.1	853	1 ENV_HV1MF	P19551 human immun
24	2685	77.3	863	1 ENV_HV1Z8	P05882 human immun
25	2678.5	77.1	859	1 ENV_HV1MA	P04583 human immun
26	2672	76.9	855	1 ENV_HV1Z6	P04580 human immun
27	2663	76.7	853	1 ENV_HV1Z2	P12487 human immun
28	2632.5	75.8	856	1 ENV_HV1ZH	P05881 human immun
29	2611.5	75.2	846	1 ENV_HV1ND	P18799 human immun
30	2595	74.7	853	1 ENV_HV1EL	P04581 human immun
31	2196	63.2	854	1 ENV_S1VCZ	P17281 chimpanzee
32	1713.5	49.3	421	1 ENV_HV1N5	P12490 human immun
33	1649.5	47.5	460	1 ENV_HV1Z3	P12491 human immun

ALIGNMENTS

RESULT 1
ENV_HV1S1

ID ENV_HV1S1 STANDARD; PRT; 847 AA.

AC P19550;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane

glycoprotein (GP120); Transmembrane glycoprotein (GP41)].

GN ENV.

OS Human immunodeficiency virus type 1 (SF162 isolate) (HIV-1).

OC Viruses; Retrovirdae; Retroviridae; Lentivirus.

OX NCBI_TaxID=11691;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90347835; PubMed=2384920;

RA Cheng-Mayer C., Quiroga M., Tung J.W., Dina D., Levy J.;

RT "Viral determinants of human immunodeficiency virus type 1 T-cell or

macrophage tropism, cytopathogenicity, and CD4 antigen modulation.";

RL J. Virol. 64:4390-4398(1990).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M65024; AAA45072.1; -.

DR PDB; LOBE; 15-MAY-97.

DR HIV; M38428; ENV\$SF162.

DR InterPro; IPR000328; Env_GP41.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal;

KW 3D-structure.

FT SIGNAL 1 29

FT CHAIN 30 502 EXTERIOR MEMBRANE GLYCOPROTEIN.

FT CHAIN 503 847 TRANSMEMBRANE GLYCOPROTEIN.

FT DISULFID 53 73 BY SIMILARITY.

FT DISULFID 118 203 BY SIMILARITY.

FT DISULFID 125 194 BY SIMILARITY.

FT DISULFID 130 155 BY SIMILARITY.

FT DISULFID 216 245 BY SIMILARITY.

FT DISULFID 226 237 BY SIMILARITY.

FT DISULFID 294 328 BY SIMILARITY.

FT DISULFID 374 435 BY SIMILARITY.

FT DISULFID 381 408 BY SIMILARITY.

FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).

Q02837 simian immu
P27977 simian immu
P18040 human immu
P27757 simian immu
P20872 human immu
P32536 human immu
P24105 human immu
P05886 simian immu
P17755 human immu
P18094 human immu
P05883 human immu
P05884 simian immu

FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	274	274	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	293	293	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	299	299	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	329	329	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	336	336	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	388	388	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	398	398	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	401	401	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	438	438	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	454	454	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	602	602	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	607	607	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	628	628	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	847 AA; 96135 MW; 0A901317FD7FF2AB CRC64;			
Query Match 98.5%; Score 3422; DB 1; Length 847;					
Best Local Similarity 95.4%; Pred. No. 1.5e-256;					
Matches 643; Conservative 0; Mismatches 3; Indels 28; Gaps 2;					
Qy	1	MRVKGIRKNYQHLWRGGTLLGLMLMICSAVEKLWVTYYYGVVWKEATTTLFCASDAKAY	60		
Db	1	MRVKGIRKNYQHLWRGGTLLGLMLMICSAVEKLWVTYYYGVVWKEATTTLFCASDAKAY	60		
Qy	61	DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNNMVEQMHEDIISLWDQSLKPCVK	120		
Db	61	DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNNMVEQMHEDIISLWDQSLKPCVK	120		
Qy	121	LTPLCVTLHCTNLKNAATNTKSSNWKEMDRGEIKNCSEFKVGAG-----	162		
Db	121	LTPLCVTLHCTNLKNAATNTKSSNWKEMDRGEIKNCSEFKVTTTSIRNKMQKEYALFYKLDVV	180		
Qy	163	-----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPCTNV	213		
Db	181	PIDNDNTSYKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPCTNV	240		
Qy	214	STVQCTHGIRPVVSTQLLNGSLAEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNN	273		
Db	241	STVQCTHGIRPVVSTQLLNGSLAEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNN	300		
Qy	274	TRKSITIGGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS	333		
Db	301	TRKSITIGGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS	360		
Qy	334	SGGDPEIVMHSFNCGGEFFYCYNSTQLFNSTWNNTIGPNNTNGTITLPCRIKQIINRWQEV	393		
Db	361	SGGDPEIVMHSFNCGGEFFYCYNSTQLFNSTWNNTIGPNNTNGTITLPCRIKQIINRWQEV	420		
Qy	394	GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKV	453		
Db	421	GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKV	480		
Qy	454	KIEPLGVAPTAKRRVWQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIV	513		
Db	481	KIEPLGVAPTAKRRVWQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIV	540		
Qy	514	QQQNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGIVGCSGKLICTTAVP	572		
Db	541	QQQNLLRAIEAQHLLQLTVWGIKQARVLAVERYLKDQQLLGIVGCSGKLICTTAVP	600		
Qy	573	WNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNOQKNEQELLELDKWASL	632		
Db	601	WNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNOQKNEQELLELDKWASL	660		
Qy	633	WNWFDISKWLWYIK 646			
Db	661	WNWFDISKWLWYIK 674			

RESULT 2					
ENV_HV1Y2					
ID	ENV_HV1Y2	STANDARD;	PRT;	843 AA.	
AC	P35961;				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].				
GN	ENV.				
OS	Human immunodeficiency virus type 1 (YU-2 isolate) (HIV-1).				
OC	Viruses; Retrovird viruses; Retroviridae; Lentivirus.				
OX	NCBI_TaxID=36377;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93021387; PubMed=1404605;				
RA	Li Y., Hui H., Burgess C.J., Price R.W., Sharp P.M., Hahn B.H., Shaw G.M.;				
RA	"Complete nucleotide sequence, genome organization, and biological properties of human immunodeficiency virus type 1 in vivo: evidence for limited defectiveness and complementation.";				
RT	J. Virol. 66:6587-6600(1992).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M93258; -; NOT_ANNOTATED_CDS.				
DR	PIR; H44001; H44001.				
DR	PDB; 1G9N; 27-DEC-00.				
DR	InterPro; IPR000328; Env_GP41.				
DR	InterPro; IPR000777; GP120.				
DR	Pfam; PF00516; GP120; 1.				
DR	Pfam; PF00517; GP41; 1.				
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal; 3D-structure.				
FT	SIGNAL	1	29		
FT	CHAIN	30	489		EXTERIOR MEMBRANE GLYCOPROTEIN.
FT	CHAIN	490	843		TRANSMEMBRANE GLYCOPROTEIN.
FT	TRANSMEM	738	755		POTENTIAL.
FT	DISULFID	53	73		BY SIMILARITY.
FT	DISULFID	118	201		BY SIMILARITY.
FT	DISULFID	125	192		BY SIMILARITY.
FT	DISULFID	130	155		BY SIMILARITY.
FT	DISULFID	214	243		BY SIMILARITY.
FT	DISULFID	224	235		BY SIMILARITY.
FT	DISULFID	292	326		BY SIMILARITY.
FT	DISULFID	373	432		BY SIMILARITY.
FT	DISULFID	380	405		BY SIMILARITY.
FT	CARBOHYD	87	87		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	129	129		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	135	135		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	138	138		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	154	154		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	158	158		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	184	184		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	193	193		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	230	230		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	237	237		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	258	258		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	272	272		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	285	285		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	291	291		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	297	297		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	327	327		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	351	351		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	381	381		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	389	389		N-LINKED (GLCNAC. . .) (POTENTIAL).

GN	ENV.				
OS	Human immunodeficiency virus type 1 (JRCSF isolate) (HIV-1).				
OC	Viruses; Retrovird viruses; Retroviridae; Lentivirus.				
OX	NCBI_TaxID=11688;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Koyanagi S., Chen I.S.Y.;				
RL	Submitted (DEC-1988) to the HIV data bank.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M38429; AAB03749.1; -.				
DR	PDB; 1CE4; 18-MAR-99.				
DR	HIV; M38429; ENV\$JRCSF.				
DR	InterPro; IPR000328; Env GP41.				
DR	InterPro; IPR000777; GP120.				
DR	Pfam; PF00516; GP120; 1.				
DR	Pfam; PF00517; GP41; 1.				
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal;				
KW	3D-structure.				
FT	SIGNAL	1	32		
FT	CHAIN	33	503		EXTERIOR MEMBRANE GLYCOPROTEIN.
FT	CHAIN	504	848		TRANSMEMBRANE GLYCOPROTEIN.
FT	DISULFID	53	73		BY SIMILARITY.
FT	DISULFID	118	203		BY SIMILARITY.
FT	DISULFID	125	194		BY SIMILARITY.
FT	DISULFID	130	154		BY SIMILARITY.
FT	DISULFID	216	245		BY SIMILARITY.
FT	DISULFID	226	237		BY SIMILARITY.
FT	DISULFID	294	328		BY SIMILARITY.
FT	DISULFID	374	437		BY SIMILARITY.
FT	DISULFID	381	410		BY SIMILARITY.
FT	CARBOHYD	87	87		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	134	134		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	137	137		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	153	153		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	157	157		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	185	185		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	195	195		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	228	228		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	239	239		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	260	260		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	274	274		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	287	287		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	293	293		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	299	299		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	329	329		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	336	336		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	352	352		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	382	382		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	388	388		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	392	392		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	403	403		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	440	440		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	453	453		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	603	603		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	608	608		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	617	617		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	629	629		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	808	808		N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	848 AA;	96475 MW;	20767F51227EC3F3	CRC64;
Query Match					
Best Local Similarity 86.4%; Score 3002.5; DB 1; Length 848;					
Matches 568; Conservative 31; Mismatches 45; Indels 33; Gaps 5;					
Qy	1	MRVKGIRKNYQHLWRGGTLLGLMLMICS	AVEKLVTVVYGVVPVWKEATTTLFCASDAKAY	60	
Db	1	MRATEIRKNYQHLWKGGTLLGLMLMICS	AAEQLVTVVYGVVPVWKEATTTLFCASDAKAY	60	
Qy	61	DTEVHNWATHACVPTDPNPQEI	VLENVTENFNMKNMVEQMHEDIISLWDQSLKPCVK	120	
Db	61	DTEVHNWATHACVPTDPNPQEV	KLENVTENFNMKNMVEQMHEDIISLWDQSLKPCVK	120	
Qy	121	LTPLCVTLHCTNLKNAATNTKSSN	WKMDRGEIKNCSEFKV-----	159	
Db	121	LTPLCVTLNCTDLRNATNTTSSW	ETMEKEIKNCSEFNITTSIRDKVQKEYALFYNLDVV	180	
Qy	160	----GAGKLINCNTSVITQACPK	VSPEPIPIHYCAPAGFAILKCN	DKFNGSGPCTNVST	215
Db	181	PIDNASYRLISCNTSVITQACPK	VSPEPIPIHYCAPAGFAILKCN	DKFNGTGPCTNVST	240
Qy	216	VQCTHGIRPVVSTQLLNGSLAE	GVVIRSENFTDNAKTIIVQLKES	VEINCTRPNNNTR	275
Db	241	VQCTHGIRPVVSTQLLNGSLAE	EEVIRSENFTNNAKTIIVQLNES	VINCTRPNNNTR	300
Qy	276	KSITIGPRAFYATGDIIGDIR	QAHCHNISGEKWNNTLKQIVTK	LQAQFG-NKTIIVFKQSS	334
Db	301	KSINIGPRALYTTGEIIGDIR	QAHCHNLSTQWENTLEQIAIK	KEQFGNKNKTIIFNPSS	360
Qy	335	GGDPEIVMHSFNCGGEFFYC	NSTQLPNSTWNNTIGPNNTNG	TITLPCRIKQIINRWQEVG	394
Db	361	GGDPEIVTHSFNCGGEFFYC	NSTQLF--TWNDRKLNNTGRN	ITLPCRIKQIINMWQEVG	418
Qy	395	KAMYAPPPIRGQIRCSSNIT	GLLLTRDGGKEISNTTEIFRP	CGGDMRDNWRSELYKYKVVK	454
Db	419	KAMYAPPPIRGQIRCSSNIT	GLLLTRDGGKD-TNGTEIFRP	CGGDMRDNWRSELYKYKVVK	477
Qy	455	IEPLGVAPTAKRRVVQREK	RAVTLGAMFLGFLGAAGSTM	GARSLTLTVQARQLLSGIVQ	514
Db	478	IEPLGVAPTAKRRVVQREK	RAVGLGALFLGFLGAAGSTM	GASITLTVQARQLLSGIVQ	537
Qy	515	QNNLLRAIEAQHLLQLTVWG	IK-LQARVLAVERYLKDQQL	LLGIWCSGKLICTTAVPW	573
Db	538	QNNLLRAIEAQHLLQLTVWG	IKQLQARVLAVERYLRDQ	QLLGIWCSGKLICTTVPW	597
Qy	574	NASWSNKSILDOIWNMT	MEWEREIDNYTNLIYTLIE	SQNOEKNEQELLELDKWASLW	633
Db	598	NTSWSNKSILNEIWDN	MTWKWEREIDNYTHIISL	IEQSNOEKNEQELLALDKWASLW	657
Qy	634	NWFDISKWLWYIK	646		
Db	658	NWFEDITKWLWYIK	670		
RESULT 3					
ENV_HV1JR					
ID	ENV_HV1JR	STANDARD;	PRT;	848 AA.	
AC	P20871;				
DT	01-FEB-1991	(Rel. 17, Created)			
DT	01-FEB-1991	(Rel. 17, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].				

Db	1	MRVKGIRKKNYQHLWKGGILLGLTLMICSAVEKLTWTVYYGVVWKEATTTTLFCASDAKAY	60
Qy	61	DTEVHNWATHACVPTDPNPQEIIVLENVTFNFMWKNMVEQMHEDIISLWDQSLKPCVK	120
Db	61	DTEVHNWATHACVPTDPNPQEVVLENVTFDFNMWKNMVEQMEDVINLWDQSLKPCVK	120
Qy	121	LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEFKVGAG	162
Db	121	LTPLCVTLNCKDV-NATNTTSSEGWMERGEIKNCSEFNITKSIRDKVQKEYALFYKLDVV	179
Qy	163	-----KLINCNTSVITQACPKVSEFEPIPIHYCAPAGFAILKCNDDKKFNGSGPCTN	212
Db	180	PIDNKNTKYRLISCNTSVITQACPKVSEFEPIPIHYCAPAGFAILKCNKTFNGKGQCKN	239
Qy	213	VSTVQCTHGIRPVVSTQLLNGLSLABEGVWIRSENFDTNNAKTIIVQLKESVEINCTRPNN	272
Db	240	VSTVQCTHGIRPVVSTQLLNGLSLABEKVVIRSDNFTDNAKTIIVQLNESVKINCTRPSN	299
Qy	273	NTRKSTIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIIVFKQ	332
Db	300	NTRKSTIHGPGRAFYTTGEIIGDIRQAHCNISRAQWNNTLKQIVEKLREQFNKTIIVFTH	359
Qy	333	SSGDPETIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNNTNG--TITLPCRICKQIINRW	390
Db	360	SSGDPETIVMHSFNCGGEFFYCNSTQLFNSTWNNTDEKSSGTEGNDTIILPCRICKQIINMW	419
Qy	391	QEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKY	450
Db	420	QEVGKAMYAPPIKGQIRCSSNITGLLLTRDGGKNESE-IEIFRPGGDMRDNRSELYKY	478
Qy	451	KVKKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLS	510
Db	479	KVKKIEPLGVAPTAKRRVVQREKRAVIGIGALFLGFLGAAGSTMGARSMTLTVQARQLLS	538
Qy	511	GIVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGWCGSGKLICTT	569
Db	539	GIVQQNNLLRAIEAQHMLQLTVWGIKQLQARVLAVERYLKDQQLMGWCGSGKLICTT	598
Qy	570	AVPWNASNKSLDQIWNNTMWEREIDNYTNLIYTLIEESNQOEKNEQELLELDKW	629
Db	599	AVPWNTSWNSKSLDSIWNNTMWEMEKEBIENYTNTIYTLIEESIQOEKNEQELLELDKW	658
Qy	630	ASLWNWFDISKWLWYIK	646
Db	659	ASLWNWFGITKWLWYIK	675
RESULT 4			
ENV_HV1SC	ID	ENV_HV1SC	STANDARD; PRT; 856 AA.
AC	P05878;		
DT	01-NOV-1988	(Rel. 09, Created)	
DT	01-NOV-1988	(Rel. 09, Last sequence update)	
DT	15-JUL-1999	(Rel. 38, Last annotation update)	
DE	Envelope polyprotein GP160 precursor [Contains: Exterior membrane		
DE	glycoprotein (GP120); Transmembrane glycoprotein (GP41)].		
GN	ENV.		
OS	Human immunodeficiency virus type 1 (SC isolate) (HIV-1).		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88219542;	PubMed=3369091;	
RA	Gurgo C., Guo H.-G., Franchini G., Aldovini A., Collalti E.,		
RA	Farrell K., Wong-Staal F., Gallo R.C., Reitz M.S. Jr.;		
RT	"Envelope sequences of two new United States HIV-1 isolates.";		
RL	Virology 164:531-536(1988).		
CC	-1- MISCELLANEOUS: THE SC ISOLATE WAS TAKEN FROM AN ARC PATIENT IN		
CC	1984 IN SOUTHERN CALIFORNIA.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		

CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M17450; -; NOT_ANNOTATED_CDS.		
DR	HIV; M17450; ENV\$SC.		
DR	InterPro; IPR000328; Env GP41.		
DR	InterPro; IPR000777; GP120.		
DR	Pfam; PF00516; GP120; 1.		
DR	Pfam; PF00517; GP41; 1.		
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;		
KW	Signal.		
FT	SIGNAL	1	29
FT	CHAIN	30	510
FT	CHAIN	511	856
FT	SITE	760	760
FT	DISULFID	53	73
FT	DISULFID	118	206
FT	DISULFID	125	197
FT	DISULFID	130	160
FT	DISULFID	219	247
FT	DISULFID	228	239
FT	DISULFID	296	330
FT	DISULFID	376	439
FT	DISULFID	383	412
FT	CARBOHYD	87	87
FT	CARBOHYD	129	129
FT	CARBOHYD	135	135
FT	CARBOHYD	140	140
FT	CARBOHYD	143	143
FT	CARBOHYD	159	159
FT	CARBOHYD	163	163
FT	CARBOHYD	189	189
FT	CARBOHYD	198	198
FT	CARBOHYD	234	234
FT	CARBOHYD	241	241
FT	CARBOHYD	262	262
FT	CARBOHYD	276	276
FT	CARBOHYD	295	295
FT	CARBOHYD	301	301
FT	CARBOHYD	302	302
FT	CARBOHYD	331	331
FT	CARBOHYD	338	338
FT	CARBOHYD	354	354
FT	CARBOHYD	360	360
FT	CARBOHYD	384	384
FT	CARBOHYD	394	394
FT	CARBOHYD	400	400
FT	CARBOHYD	405	405
FT	CARBOHYD	442	442
FT	CARBOHYD	457	457
FT	CARBOHYD	611	611
FT	CARBOHYD	616	616
FT	CARBOHYD	625	625
FT	CARBOHYD	637	637
FT	CARBOHYD	674	674
FT	CARBOHYD	816	816
SQ	SEQUENCE	856 AA; 97055 MW; DAF4DA600EBA7A08	CRC64;

Query Match			
Best Local Similarity		86.3%;	Score 2997.5; DB 1; Length 856;
Matches 568;		Conservative 34;	Mismatches 43; Indels 39; Gaps 7;

Qy	1	MRVKGIRKKNYQHLWRGGTLLGLMLMICSASAVEKLTWTVYYGVVWKEATTTTLFCASDAKAY	60
Db	1	MRVKGSGRNYQHLWRWGTMLLGILMICSAAEQLTWTVYYGVVWKEATTTTLFCASDAKAY	60
Qy	61	DTEVHNWATHACVPTDPNPQEIIVLENVTFNFMWKNMVEQMHEDIISLWDQSLKPCVK	120
Db	61	DTEVHNWATHACVPTDPNPQEVVLENVTFNFMWKNMVEQMHEDIISLWDQSLKPCVK	120

```
QY 121 LTPLCVTLHCTNLK-----NATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db 121 LTPLCVTLNCTNLNDTSTNATNTTSNRGKMEGEMTNCSEFNITTSIRSKVQKEYALFY 180
QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKSGGPC 210
Db 181 KLDVVPIDNTSYTLINCNTSVITQACPKVSFEPIPIHYCA-RWFAILNCNNKFNKSGGPC 239
QY 211 TNVSTVQCTHGIRPVVSTQLLLNGSLABEGVVIRSENFNTDNAKTIIVQLKESVEINCTRP 270
Db 240 TNVSTVQCTHGIRPVVSTHLLNGSLABEEVVLRSSENFNTDNAKTIIVQLKEAVEINCTRP 299
QY 271 NNNTKRSITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIIV 330
Db 300 NNNTTRSIIHGPGRAFYATGDIIGDIRQAHCNISRAKWNNTLKQIVIKLRDQFENKTIIF 359
QY 331 KQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNNTG--TITLPCRICKIIN 388
Db 360 NRSSGGDPEIVMHSFNCGGEFFYCNSTQLFSSTWNGTEGSSNNTGGNDTITLPCRICKIIN 419
QY 389 RWOEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGKEISNT----TEIFRPGGDMRDNR 444
Db 420 MWQEVGKAMYAPPIKGQVKCSSNITGLLLTRDGGNSKNGSKNENTEIFRPGGDMRDNR 479
QY 445 SELYKYKVVKIEPLGVAPTAKRRVVQREKRAV-TLGAMFLGFLGAAGSTMGARSLTLTV 503
Db 480 SELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGTIGAMFLGFLGAAGSTMGATSMTLTV 539
QY 504 QARQLLSGIVQQNNLLRAIEAQOHLQLTVWGIK-LQARVLAVERYLKDQQLLGWGS 562
Db 540 QARLLLSGIVQQNNLLRAIEAQOHLQLTVWGIKQLQARVLAVERYLRDQQLLGWGS 599
QY 563 GKLICTTAVPWNASWSNKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQE 622
Db 600 GKLICTTTVPWNTSWNSKSLDKIWGNNTWMEWEREIDNYTSLIYTLIEESQNQOEKNEQE 659
QY 623 LLELDKWASLWNWFDISKWLWYIK 646
Db 660 LLELDKWASLWNWFENITNLWLYIK 683

RESULT 5
ENV_HV1A2
ID ENV HV1A2 STANDARD; PRT; 855 AA.
AC P03378;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
DE glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (ARV2/SF2 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11685;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85090453; PubMed=2578227;
RA Sanchez-Pescador R., Power M.D., Barr P.J., Steimer K.S.,
RA Stempien M.M., Brown-Shimer S.L., Gee W.W., Renard A., Randolph A.,
RA Levy J.A., Dina D., Luciw P.A.;
RT "Nucleotide sequence and expression of an AIDS-associated retrovirus
RT (ARV-2).";
RL Science 227:484-492(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02007; AAB59882.1; -.
```

```
DR PIR; A03976; VCLJA2.
DR HIV; K02007; ENVSSF2.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW signal.
FT SIGNAL 1 29
FT CHAIN 30 509
FT CHAIN 510 855
FT DISULFID 53 73
FT DISULFID 118 208
FT DISULFID 125 199
FT DISULFID 130 155
FT DISULFID 221 250
FT DISULFID 231 242
FT DISULFID 299 333
FT DISULFID 380 442
FT DISULFID 387 415
FT CARBOHYD 87 87
FT CARBOHYD 129 129
FT CARBOHYD 140 140
FT CARBOHYD 154 154
FT CARBOHYD 158 158
FT CARBOHYD 184 184
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 233 233
FT CARBOHYD 244 244
FT CARBOHYD 265 265
FT CARBOHYD 279 279
FT CARBOHYD 292 292
FT CARBOHYD 298 298
FT CARBOHYD 304 304
FT CARBOHYD 334 334
FT CARBOHYD 341 341
FT CARBOHYD 358 358
FT CARBOHYD 364 364
FT CARBOHYD 388 388
FT CARBOHYD 394 394
FT CARBOHYD 400 400
FT CARBOHYD 408 408
FT CARBOHYD 445 445
FT CARBOHYD 458 458
FT CARBOHYD 461 461
FT CARBOHYD 610 610
FT CARBOHYD 615 615
FT CARBOHYD 624 624
FT CARBOHYD 636 636
FT CARBOHYD 815 815
SQ SEQUENCE 855 AA; 97438 MW; A3BC20573AAC41A2 CRC64;

Query Match 85.9%; Score 2984; DB 1; Length 855;
Best Local Similarity 82.4%; Pred. No. 1.le-222;
Matches 563; Conservative 34; Mismatches 48; Indels 38; Gaps 6;

QY 1 MRVKGIRKNYQHLWRGGTLLGLMLMICS AVEKLVWTVVYGVVWKEATTLFCASDAKAY 60
Db 1 MKVKGTRRNYQHLWRWGTLLGLMLMICS ATEKLVWTVVYGVVWKEATTLFCASDARAY 60
QY 61 DTEVHNVWATHACVPTDPNPQBEIVLENTENFNMWKNNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNVWATHACVPTDPNPQBEIVLENTENFNMWKNNMVEQMHEDIISLWDQSLKPCVK 120
QY 121 LTPLCVTLHCTNLKNAATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db 121 LTPLCVTLNCTDLGKATNTNSSNWKEIKGEIKNCSFNITTSIRDKIQENALFRNLDDV 180
QY 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFNKSG 208
Db 181 PIDNASTTNTYNTYRLIHCNRSVITQACPKVSFEPIPIHYCTPAGFAILKCNKTFNGKG 240
```


Db 419 NMQEVGKAMYAPPIRGQIKCSSNITGLLLTRDGGNNNETTDEIFRPGGNNMRDNWRSE 478

Qy 447 LYKYKVVKIEPLGVAPTAKRRVVOREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQAR 506

Db 479 LYKYKVVKIEPLGVAPTAKRRVVOREKRAVGLGALFLGFLGAAGSTMGAASLTTLTVQAR 538

Qy 507 QLLSGIVQQQNNLLRAIEAQHLLQLTWGIK-LQARVLAVERYLKDQQLLGIWCSGKL 565

Db 539 LLLSGIVQQQNNLLMAIEAQHMLTLTWGIKQLQARVLAVERYLKDQQLLGIWCSGKL 598

Qy 566 ICTTAVPWNASWSNKSLSQIWNMTWMEWEREIDNTYTNLIYLIBESONQOEKNEQELLE 625

Db 599 ICTTAVPWNASWSNKSLSQIWDNMTWMEWEREIDNTYTNLIYSLIEDSQIQEKNKEKELLE 658

Qy 626 LDKWASLWNWFDISKWLWYIK 646

Db 659 LDKWASLWNWENITNWLWYIK 679

RESULT 8

ENV_HV1B1

ID ENV_HV1B1 STANDARD; PRT; 856 AA.

AC P03375;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].

GN ENV.

OS Human immunodeficiency virus type 1 (BH10 isolate) (HIV-1).

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11678;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85111123; PubMed=2578615;

RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R., Josephs S.F., Doran E.R., Rafalski J.A., Whitehorn E.A., Baumeister K., Ivanoff L., Petteway S.R. Jr., Pearson M.L., Lautenberger J.A., Papas T.S., Grayeb J., Chang N.T., Gallo R.C., Wong-Staal F.;

RA "Complete nucleotide sequence of the AIDS virus, HTLV-III.";

RT Nature 313:277-284(1985).

RL [2]

RN

RP DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.

RX MEDLINE=90285159; PubMed=2355006;

RA Leonard C.K., Spellman M.W., Riddle L., Harris R.J., Thomas J.N., Gregory T.J.;

RA "Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells.";

RL J. Biol. Chem. 265:10373-10382(1990).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M15654; AAA44205.1; -.

DR PIR; A03973; VCLJH3.

DR HIV; M15654; ENV\$BH102.

DR InterPro; IPR000328; Env GP41.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal.

KW SIGNAL

FT CHAIN 1 30

FT CHAIN 31 511

FT CHAIN 512 856

EXTERIOR MEMBRANE GLYCOPROTEIN.

TRANSMEMBRANE GLYCOPROTEIN.

FT DISULFID 54 74

FT DISULFID 119 205

FT DISULFID 126 196

FT DISULFID 131 157

FT DISULFID 218 247

FT DISULFID 228 239

FT DISULFID 296 331

FT DISULFID 378 445

FT DISULFID 385 418

FT CARBOHYD 88

FT CARBOHYD 136 136

FT CARBOHYD 141 141

FT CARBOHYD 156 156

FT CARBOHYD 160 160

FT CARBOHYD 186 186

FT CARBOHYD 197 197

FT CARBOHYD 230 230

FT CARBOHYD 234 234

FT CARBOHYD 241 241

FT CARBOHYD 262 262

FT CARBOHYD 276 276

FT CARBOHYD 289 289

FT CARBOHYD 295 295

FT CARBOHYD 301 301

FT CARBOHYD 332 332

FT CARBOHYD 339 339

FT CARBOHYD 356 356

FT CARBOHYD 386 386

FT CARBOHYD 392 392

FT CARBOHYD 397 397

FT CARBOHYD 406 406

FT CARBOHYD 448 448

FT CARBOHYD 463 463

FT CARBOHYD 611 611

FT CARBOHYD 616 616

FT CARBOHYD 625 625

FT CARBOHYD 637 637

FT CARBOHYD 674 674

FT CARBOHYD 750 750

FT CARBOHYD 816 816

SQ SEQUENCE 856 AA; 97224 MW; 0BFFB1A18931BB27 CRC64;

Query Match 84.6%; Score 2939.5; DB 1; Length 856;

Best Local Similarity 82.6%; Pred. No. 3e-219;

Matches 568; Conservative 34; Mismatches 39; Indels 47; Gaps 12;

Qy 1 MRVKGIRKKNYQHL-----WRGGTLLGLMLMICS AVEKLVWTVVYGVVWKEATTLFCASD 56

Db 1 MRVK---EKYQHLWRGWGRWGTMLGLMLMICSATEKLVWTVVYGVVWKEATTLFCASD 57

Qy 57 AKAYDTEVHNWATHACVPTDPNPQEI VLENTENFNWKNWVQMHEDIISLWDQSLK 116

Db 58 AKAYDTEVHNWATHACVPTDPNPQEI VLVNVTENFNWKNWVQMHEDIISLWDQSLK 117

Qy 117 PCVKLTPLCVTLHCTNLKNATNTKSSNWKE-MDRGEIKNCSFKVGA---GK----- 163

Db 118 PCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKEIKNCSFNISTSI RGVQKEYAFFY 177

Qy 164 -----LINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNKFN GSG 208

Db 178 KLDIIPIDNDTTSYTLTSCNTSVITQACPKVSFEPIPIHYCAPAG FAILKCNKFTNGTG 237

Qy 209 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGWIRSENFTDNAKTIIV QLKESVEINCT 268

Db 238 PCTNVSTVQCTHGIRPVVSTQLLNGSLAEEVWIRSANFTDNAKTI IVQLNQSVINCT 297

Qy 269 RPNNTRKSIITI--GPGRAFYATGDIIGDIRQAHNCNISGEKNNT LKQIVTKLQAFG-N 325

Db 298 RPNNTRKSIIRIQRPGPRAFTVIGK-IGNMRQAHNCNISRAKNW NTLKQIDSKLREQFGNN 356

Qy 326 KTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTOLFNSTWNN---- TIGPNNTNG--TITL 379

Db 357 KTIIFKQSSGGDPEIVTHSFNCGGGEFFYCNSTOLFNSTWFSN TWFNSTWSTKGSNNTGSDTITL 416

Qy 377 -----ITLPCRKIQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTT 429
Ddb 418 QKDTGDIITLPCRKIQIINRWQVVGKAMYALPIKGLIRCSSNITGLLLTRDGGGE-NQTT 476
Qy 430 E1FRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAV-TLGAMFLGFLG 488
Ddb 477 E1FRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGLGAMFLGFLG 536
Qy 489 AAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGK-LQARVLAVE 547
Ddb 537 AAGSTMGATSMALTVQARQLLSGIVQQNNLLRAIKAQOHLQLTVWGKQLQARILAVE 596
Qy 548 RYLKDOQLLGIWCGSKLICITAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYT 607
Ddb 597 RYLKDOQLLGFWCGSKLICITAVPWNASWSNKTLDQIWNMTWMEWDREIDNYTHLIYT 656
Qy 608 LIEESQOQEKNEQELLELDKASLWNWFDISKWLWYIK 646
Ddb 657 LIEESQOQEKNOQELLQDQKASLWTWSDITKWLWYIK 695

RESULT 10
ENV_HV1J3
ID ENV_HV1J3 STANDARD; PRT; 867 AA.
AC p12489;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (JH3 isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11694;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=89352108; PubMed=2669897;
RA Komiyama N., Hattori N., Inoue J., Sakuma S., Kurimura T., Yoshida M.;
RT "Nucleotide sequences of gag and env genes of a Japanese isolate of
RT HIV-1 and their expression in bacteria."
RL AIDS Res. Hum. Retroviruses 5:411-419(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21138; AAB03526.1; --
DR HIV; M21138; ENV\$JH3.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
FT Signal.
FT SIGNAL 1 30
FT CHAIN 31 516 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 517 867 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 217 BY SIMILARITY.
FT DISULFID 125 208 BY SIMILARITY.
FT DISULFID 130 160 BY SIMILARITY.
FT DISULFID 230 259 BY SIMILARITY.
FT DISULFID 240 251 BY SIMILARITY.
FT DISULFID 308 342 BY SIMILARITY.
FT DISULFID 388 457 BY SIMILARITY.
FT DISULFID 395 430 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 307 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 396 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 622 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 648 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 867 AA; 98399 MW; 5F2310146B8E8680 CRC64;
Query Match 84.5%; Score 2935; DB 1; Length 867;
Best Local Similarity 80.2%; Pred. No. 6.9e-219;
Matches 558; Conservative 38; Mismatches 48; Indels 52; Gaps 7;
Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLICSAVEKLVWTVVYGVVWKEATTLFCASDAKAY 60
Ddb 1 MRVKGIRKNYQHLWRWGTMLLGLMLICSAEQLWTVVYGVVWKEATTLFCASDAKAY 60
Qy 61 DTEVHNVWATHACVPTDPNPQEIIVLENTENFMWKNMVEQMHEDIISLWDQSLKPCVK 120
Ddb 61 DTEVHNVWATHACVPTDPNPQEVVLENTKFNWKNMVEQMHEDIISLWDQSLKPCVK 120
Qy 121 LTPLCVTLHCTN----LKNATNTKSSNWKEMDRGEIKNSFKVGAG----- 162
Ddb 121 LTPLCVTLNCIDWGNDTSPNATNTTSSGGEKMEKGMKNSFNITTSIRDKVQKEHALFY 180
Qy 163 -----KLINCNTSVITQACPKVSFEPPIHYCAPAGFAILKC 199
Ddb 181 KHDVVPINNSTKDNKNDNSTRYRLISCNTSVITQACPKISFEPPIHYCAPAGFAIIC 240
Qy 200 NDKKFNKSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFTDAKTIIVQL 259
Ddb 241 NDKKFNKSGPCTNVSTVQCTHGIRKPVVSTQLLNGSLAEVEVIRSENFTDAKTIIVQL 300
Qy 260 KESVEINCTRPNNNTRKSIITIGPGRAPYATGDIIGDIRQAHCHNISGEKWNNTLKQIVTKL 319
Ddb 301 KEPVVINCTRPSTKTRRRRIHIGPGRAPYTTKQIAGDLRQAHCHNRRARNATLKQIVGKL 360
Qy 320 QAQFGNKTIVFKOSSGGDPEIVMHSFNCGGEFFYCNSTQLF-----NSTWNNTIGPNNT 373
Ddb 361 RKQFVNKTIVFNRSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSWLSNSTWNDETSNNT 420
Qy 374 NG--TITLPCRKIQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEI 431
Ddb 421 GGNDTITLPCRKIQIINRWQEVGKAMYAPPPIEGQIRCSSNITGLLLTRDGG-DNQNETET 479
Qy 432 FRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAG 491
Ddb 480 FRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGIVGFLGFLGAAG 539
Qy 492 STMGARSLLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTVWGK-LQARVLAVERYL 550
Ddb 540 STMGA-SMTLTVQARLLLSGIVQQNNLLRAIEGQQOHLQLTVWGKIQARILAVERYL 598

RP SEQUENCE FROM N.A.
RX MEDLINE=85111123; PubMed=2578615;
RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R.,
RA Josephs S.F., Doran E.R., Rafalski J.A., Whitehorn E.A.,
RA Baumeister K., Ivanoff L., Petteway S.R. Jr., Pearson M.L.,
RA Lautenberger J.A., Papas T.S., Ghrayeb J., Chang N.T., Gallo R.C.,
RA Wong-Staal F.;
RT "Complete nucleotide sequence of the AIDS virus, HTLV-III.";
RL Nature 313:277-284(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02011; AAA44661.1; -.
DR PDB; 1DDH; 13-JAN-99.
DR PDB; 1HHG; 31-OCT-93.
DR PDB; 1QO3; 02-JAN-00.
DR PDB; 1SZT; 24-DEC-97.
DR HIV; K02011; ENVSBH8.
DR GlycoSuiteDB; P04582; -.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal;
KW 3D-structure.
FT SIGNAL 1 30
FT CHAIN 31 506 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 507 851 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 54 74 BY SIMILARITY.
FT DISULFID 119 205 BY SIMILARITY.
FT DISULFID 126 196 BY SIMILARITY.
FT DISULFID 131 157 BY SIMILARITY.
FT DISULFID 218 247 BY SIMILARITY.
FT DISULFID 228 239 BY SIMILARITY.
FT DISULFID 296 331 BY SIMILARITY.
FT DISULFID 378 440 BY SIMILARITY.
FT DISULFID 385 413 BY SIMILARITY.
FT CARBOHYD 88 136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 160 160 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 166 186 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 606 606 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 632 632 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 745 745 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 851 AA; 96644 MW; D16A3C90857785F1 CRC64;

Query Match		84.0%;	Score 2918;	DB 1;	Length 851;		
Best Local Similarity		82.2%;	Pred. No. 1.4e-217;				
Matches 562;		Conservative 35;	Mismatches 43;	Indels 44;	Gaps 11;		
Qy	1	MRVKGIRKNYQHL	---WRGGTLLGLMLMICS	AVEKLVWTVVYGVVPWKEATTTLFCASD	56		
Db	1	MRVK--EKYQHLWRWGRWG	TLLGLMLMICS	ATEKLVWTVVYGVVPWKEATTTLFCASD	57		
Qy	57	AKAYDEVHNVWATHACVPTDPNPQEI	VLENVTENFNMWKNMVEQMHE	DIISLWDQSLK	116		
Db	58	AKAYDEVHNVWATHACVPTDPNPQEV	VLNVTFENFNMWKNMVEQMHE	DIISLWDQSLK	117		
Qy	117	PCVKLTPLCVTLHCTNLKNA	TNWKSSNWE-MDRGEIKNC	SPKVGAGK-----	163		
Db	118	PCVKLTPLCVSLKCTDLKND	TNTNSSGRMIMEKEIKNC	SFNISTSKRGKVQKEYAFFY	177		
Qy	164	-----	LINCNTSVITQACPKVS	FEPIPIHYCAPAGPAILK	CNDKKFNGSG	208	
Db	178	KLDIIPIDNDTTSYTLT	SCNTSVITQACPKVS	FEPIPIHYCAPAGPAILK	CNNKTFNGTG	237	
Qy	209	PCTNVSTVQCTHGIRPV	VSTQLLNGSLAE	EGVVRSENF	TNAKTIIVQLKES	VEINCT	268
Db	238	PCTNVSTVQCTHGIRPV	VSTQLLNGSLAE	EEVVRSENF	TNAKTIIVQLDTS	VEINCT	297
Qy	269	RPNNTRKSITI--	GPGRAFATGDIIGDIRQAHC	NISGEKWNNTLKQIVTK	LQAOFG-N	325	
Db	298	RPNNTRKKIRIORG	PGRAFVTIGK-IGNMRQAHC	NISRAKWNATLKQIDSK	LRQOFGNN	356	
Qy	326	KTIVFKOSSGGDPEI	VMHSFNCGGEFFYC	NSTQLFNSTWNNTIGPN	NTNG--TITLPCRI	383	
Db	357	KTIIFKOSSGGDPEI	VTHSFNCGGEFFYC	NSTQLFNSTW-STKGS	NNTEGSDTITLPCRI	415	
Qy	384	KQIINRWQEVGKAMY	APPPIRGQIRC	SSNITGLLLTRD	GGKEISNTTEIFRPGG	DMRDNW	443
Db	416	KQIINMWQEVGKAMY	APPPISGQIRC	SSNITGLLLTRD	GGNS-NNSEIFRPGG	DMRDNW	474
Qy	444	RSELYKYKVVKIE	PLGVAPTAKRRVVQRE	KRAVTLGMFLGLGA	AGSTMGARSLTLTV	503	
Db	475	RSELYKYKVVKIE	PLGVAPTAKRRVVQRE	KRAVGIGALFLG	FLGAAGSTMGAASMTLTV	534	
Qy	504	QARQLLSGIVQQ	NNLLRAIEAQOHL	LQLTWGIK-LQARVLA	VERYLKDQQLLGI	WGCS	562
Db	535	QARQLLSGIVQQ	NNLLRAIEGQOHL	LQLTWGIKQ	LQARILAVERYLKDQQLLGI	WGCS	594
Qy	563	GKLICTTAVP	PNASWSNKS	LQIWNNTWME	WEREIDNYTNLIYTLIE	SQOQEKNEQ	622
Db	595	GKLICTTAVP	PNASWSNKS	LQIWNNTWME	WDREINNYTSLI	HSLSIESQOQEKNEQ	654
Qy	623	LLELDKWASL	NWFEDISK	WLYIK	646		
Db	655	LLELDKWASL	NWFENITN	WLYIK	678		

Search completed: February 25, 2004, 14:16:59
Job time : 13.2348 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 14:12:26 ; Search time 40.8538 Seconds
(without alignments)
4989.132 Million cell updates/sec

Title: US-09-891-609A-2
Perfect score: 3474
Sequence: 1 MRVKGIRKKNYQHLWRGGTLL.....DKWASLWNWFDISKWLWYIK 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_irvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3323	95.7	847	15 Q8AR21	Q8ar21 human immun
2	3096.5	89.1	850	15 Q7SVL4	Q7svl4 human immun
3	3093.5	89.0	854	15 Q7SVL3	Q7svl3 human immun
4	3092.5	89.0	860	15 Q7SVL7	Q7svl7 human immun
5	3079.5	88.6	852	15 Q7SVL5	Q7svl5 human immun
6	3074	88.5	853	15 Q7SVL6	Q7svl6 human immun
7	3049.5	87.8	854	15 O40222	O40222 human immun
8	3040.5	87.5	846	15 Q9PXW7	Q9pxw7 human immun
9	3035	87.4	847	15 Q7SVL60	Q7svl60 human immun
10	3030	87.2	843	15 Q70150	Q70150 human immun
11	3025.5	87.1	852	15 O41883	O41883 human immun
12	3022.5	87.0	852	15 O92761	O92761 human immun
13	3022	87.0	853	15 O56108	O56108 human immun
14	3020	86.9	851	15 O56110	O56110 human immun
15	3017	86.8	845	15 O70679	O70679 human immun
16	3017	86.8	853	15 Q03811	Q03811 human immun

17	3010.5	86.7	858	15 Q7SVL1	Q7svl1 human immun
18	3009.5	86.6	850	15 Q74812	Q74812 human immun
19	3003.5	86.5	856	15 Q77694	Q77694 human immun
20	3002.5	86.4	856	15 Q72993	Q72993 human immun
21	3001.5	86.4	854	15 Q78225	Q78225 human immun
22	3000	86.4	853	15 Q7ZJP5	Q7zjf5 human immun
23	3000	86.4	859	15 Q7SVL2	Q7svl2 human immun
24	2999.5	86.3	864	15 Q8UL53	Q8ul53 human immun
25	2997	86.3	859	15 Q7SUT0	Q7sut0 human immun
26	2992.5	86.1	848	15 Q74999	Q74999 human immun
27	2992.5	86.1	852	15 Q41885	O41885 human immun
28	2992.5	86.1	860	15 Q7ZJD7	Q7zjd7 human immun
29	2992	86.1	859	15 Q7SUS9	Q7sus9 human immun
30	2991	86.1	839	15 Q73364	Q73364 human immun
31	2990	86.1	861	15 O56109	O56109 human immun
32	2988.5	86.0	850	15 O11946	O11946 human immun
33	2988	86.0	855	15 Q7SVL0	Q7svl0 human immun
34	2987	86.0	849	15 Q77368	Q77368 human immun
35	2986	86.0	861	15 Q9IUY9	Q9iuy9 human immun
36	2985	85.9	859	15 O71260	O71260 human immun
37	2984.5	85.9	854	15 O92762	O92762 human immun
38	2983	85.9	855	15 Q8Q367	Q8q367 human immun
39	2982.5	85.9	850	15 O56564	O56564 human immun
40	2982.5	85.9	868	15 Q9WJW5	Q9wjw5 human immun
41	2981	85.8	863	15 O42031	O42031 human immun
42	2979	85.8	859	15 Q7SUT4	Q7sut4 human immun
43	2978.5	85.7	855	15 Q03805	Q03805 human immun
44	2978	85.7	863	15 Q9WJU8	Q9wju8 human immun
45	2977.5	85.7	864	15 Q9YP39	Q9yp39 human immun

ALIGNMENTS

RESULT 1

Q8AR21 ID Q8AR21 PRELIMINARY; PRT; 847 AA.
AC Q8AR21;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHIVSF162P3.2;
RA Gao F.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF536757; AAN05642.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
SQ SEQUENCE 847 AA; 95879 MW; 9CF863B97B3BB54C CRC64;

Query Match 95.7%; Score 3323; DB 15; Length 847;
Best Local Similarity 92.4%; Pred. No. 8.3e-264;
Matches 623; Conservative 7; Mismatches 16; Indels 28; Gaps 2;

Qy	1	MRVKGIRKKNYQHLWRGGTLLGLMLICSAVEKLWVTYYGVVWKEATTTLFCASDAKAY 60	
Db	1	MRVKGIRKKNYQHLWRGGTLLGLMLICSAVEKLWVTYYGVVWKEATTTLFCASDAKAY 60	
Qy	61	DTEVHNWATHACVPTDPNPQEIIVLENTENFNWKNMVEQMHEDIISLWDSLKPCVK 120	
Db	61	DTEVHNWATHACVPTDPNPQEIIVLENTENFNWKNMVEQMHEDIISLWDSLKPCVK 120	
Qy	121	LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEFKVGAG----- 162	

Db 121 LTPLCVTLHCTNLNENATNTSSNWKEMNRGEIKNCSEFNVTTSIGNKMQKEYALFYRLDVV 180
Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNV 213
Db 181 PIDNDNTSYNLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCINV 240
Qy 214 STVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNN 273
Db 241 STVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNKTIIIVQLKESVEINCTRPNN 300
Qy 274 TRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQS 333
Db 301 TRKSIPIGPKAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFENKTIVFKQS 360
Qy 334 SGGDPEIVMHSFNCGGEFFCYCNSTQLFNSTWNNTIGPNNTNGTITLPCRICKQIINRWQEV 393
Db 361 SGGDPEIVMHSFNCGGEFFCYCNSTQLFNSTWNNTIGPNNTNGTITLPCRICKQIINRWQEV 420
Qy 394 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKV 453
Db 421 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGREVGNTTEIFRPGGDMRDNRSELYKYKV 480
Qy 454 KIEPLGVAPTAKARRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIV 513
Db 481 KIEPLGVAPTAKARRVVQREKRAVTLGAVFLGFLGAAGSTMGAASLTLTVQARQLLSGIV 540
Qy 514 QQQNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLGIWCGSGKLICTTAVP 572
Db 541 QQQNLLRAIEAQORLLQLTVWGIKQARVLAVERYLKDQQLGIWCGSGKLICTTAVP 600
Qy 573 WNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASL 632
Db 601 WNASWSNKSLDQIWNMTWMEWEREIGNYTNLIYTLIEESNQOEKNEQELLELDKWASL 660
Qy 633 WNWFDISKWLWYIK 646
Db 661 WNWLDISKWLWYIK 674

RESULT 2
Q7SVL4
ID Q7SVL4 PRELIMINARY; PRT; 850 AA.
AC Q7SVL4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81NJ;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247221; AAP37149.1; -.
KW Envelope protein.
SQ SEQUENCE 850 AA; 96583 MW; 62ED5F6AB033D20B CRC64;

Query Match 89.1%; Score 3096.5; DB 15; Length 850;
Best Local Similarity 86.1%; Pred. No. 3.3e-245;
Matches 584; Conservative 26; Mismatches 35; Indels 33; Gaps 4;
Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICSABEKLWVTYYGVVPVWKEATTTLFCASDAKAY 60
Db 1 MRVKEIRKNYQHLWRWGTMLLGLMILMICSABEKLWVTYYGVVPVWKEATTTLFCASDAKAY 60

Qy 61 DTEVHNWATHACVPTDPNPQEIIVLENVTENFNMKNNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWATHACVPTDPNPQEWLENVTENFNMKNNMVEQMHEDIISLWDQSLKPCVK 120
Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSEFKVAG----- 162
Db 121 LTPLCVTLNCTDLRNATNTSSGGTMERGEIKNCSEFNITTSIRDKVQKEYALFYKLDVV 180
Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCT 211
Db 181 PIDNDNTTTSYRLISCNTSVITQACPKVSFEPIPIHYCTPAGFAILCKDKKFNGTGPCT 240
Qy 212 NVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPN 271
Db 241 NVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSDNFTDNAKTIIVQLKESVEINCTRPN 300
Qy 272 NNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFK 331
Db 301 NNTRKSIHIGPGRAFYTTGTGIIIGDIRQAHCNLSRAKWDNTLKQIVRKLREQFGNKTIVEN 360
Qy 332 QSSGGDPEIVMHSFNCGGEFFCYCNSTQLFNSTWNNTIGPNNTNG--TITLPCRICKQIINR 389
Db 361 QSSGGDPEIVTHSFNCGGEFFCYDSTQLFNSTWNVTGEGSNNTGNTITLTLPCRICKQIINM 420
Qy 390 WQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYK 449
Db 421 WQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGNESE-TEIFRPGGDMRDNRSELYK 479
Qy 450 YKVVKIEPLGVAPTAKARRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLL 509
Db 480 YKVVKIEPLGVAPTAKARRVVQREKRAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLL 539
Qy 510 SGIVQQQNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLGIWCGSGKLICT 568
Db 540 SGIVQQQNLLRAIEAQHLLQLTVWGIKQOLQARVLAVERYLKDQQLGIWCGSGKLICT 599
Qy 569 TAVPWNASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLELDK 628
Db 600 TTVPWNASWSNKSLDQIWNMTWMEWEREINNYTSLIYTLIEESNQOEKNEQELLELDK 659
Qy 629 WASLWNWFDISKWLWYIK 646
Db 660 WASLWNWFDITKWLWYIK 677

RESULT 3
Q7SVL3
ID Q7SVL3 PRELIMINARY; PRT; 854 AA.
AC Q7SVL3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81NY1;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247222; AAP37150.1; -.
KW Envelope protein.
SQ SEQUENCE 854 AA; 97048 MW; 36379231FF3AF12C CRC64;

Query Match 89.0%; Score 3093.5; DB 15; Length 854;
Best Local Similarity 85.8%; Pred. No. 5.9e-245;
Matches 584; Conservative 22; Mismatches 40; Indels 35; Gaps 5;

RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247220; AAP37148.1; -.
KW Envelope protein.
SQ SEQUENCE 852 AA; 96680 MW; F8D811028025D998 CRC64;

Query Match 88.6%; Score 3079.5; DB 15; Length 852;
Best Local Similarity 85.3%; Pred. No. 8.3e-244;
Matches 580; Conservative 27; Mismatches 38; Indels 35; Gaps 4;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICS AVEKLVWTVVYGVVPVWKEATTTLFCASDAKAY 60
Db 1 MRVKEIGKNYQYLRWGTMLLGLMICS AAEQLWTVVYGVVPVWKEATTTLFCASDAKAY 60

Qy 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWATHACVPTDPNPQEV GLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNC SFKVGAG----- 162
Db 121 LTPLCVTLNCTDLGNTTNTSSGGMERGEIKNC SFNITTSIRDKVQKEYALLYKLDIV 180

Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNV 213
Db 181 PIDDNNTSYRLISCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKFNGKGPCTNV 240

Qy 214 STVQCTHGIRPVVSTQLLLNGSLAE EGVVIRSENFDTNAKTIIVQLKESVEINCTRPNN 273
Db 241 STVQCTHGIRPVVSTQLLLNGSLAE EGVVIRSDNFTNAKTIIVQLKESVEINCTRPNN 300

Qy 274 TRKSIITGPGRAFATGDIIGDIRQAH CNISGEKWNNTLKQIVTKLQAQFGNKTIIVFKS 333
Db 301 TRKSIHIGPGRAFYTGTGDIVGDIRQAH CNISKAKWNTTLKQIVTKLREQFGNRTIVFNQS 360

Qy 334 SGGDPEIVMHSFNCGGEFFCYNSTQLFN-----STWNNTIGPNTNGTITLPCRICKQII 387
Db 361 SGGDPEIVMHSFNCGGEFFCYNSTQLFNSTWNASSTWNDTEGSNNTEGTITLPCRICKQII 420

Qy 388 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 447
Db 421 NMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGSN-ENDTEIFRPGGDMRDNRSEL 479

Qy 448 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 507
Db 480 YKYKVVKIEPLGVAPTAKRRVVQREKRAVIGAVFLGFLGAAGSTMGAASMTLTVQARL 539

Qy 508 LLSGIVQQNNLLRAIEAQOHLQLTVWG IK-LQARVLAVERYLKDQQLLGWCSGKLI 566
Db 540 LLSGIVQQNNLLRAIEAQOHLQLTVWG IKQLQARVLAVERYLKDQQLLGWCSGKLI 599

Qy 567 CTTAVPWNASWSNKS LDIWNNMTWMEWEREIDNYTNLIYTLIEESQNQQEKNEQELLEL 626
Db 600 CTTTVPWNASWSNKS LDIWNNMTWMEWDREINNYTSLIYTLIEESQNQQEKNEQELLEL 659

Qy 627 DKWASLWNWFDISKWLWYIK 646
Db 660 DKWASLWNWFDTIKWLWYIK 679

RESULT 6
Q7SVL6
ID Q7SVL6 PRELIMINARY; PRT; 853 AA.
AC Q7SVL6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=81CA2;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247219; AAP37147.1; -.
KW Envelope protein.
SQ SEQUENCE 853 AA; 96545 MW; E3B2830A1261E237 CRC64;

Query Match 88.5%; Score 3074; DB 15; Length 853;
Best Local Similarity 84.9%; Pred. No. 2.3e-243;
Matches 580; Conservative 32; Mismatches 31; Indels 40; Gaps 5;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICS AVEKLVWTVVYGVVPVWKEATTTLFCASDAKAY 60
Db 1 MRVKGIRKNYQHLWRWGTMLLGLMICS ATEKLVWTVVYGVVPVWKEATTTLFCASDAKAY 60

Qy 61 DTEVHNWATHACVPTDPNPQEI VLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWATHACVPTDPNPQEV VLGNTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNC SFKVGAG----- 162
Db 121 LTPLCVTLNCTDLRNATNTSSGEKMEGGEIKNC SFNITTSLRDKMQKEYALFYKLDVV 180

Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNDKKENGSGPCT 211
Db 181 PIDDNNTSYRLISCNTSVITQACPKVSFEPIPIHFCAAGFAILKCNDKKFSKGPCT 240

Qy 212 NVSTVQCTHGIRPVVSTQLLLNGSLAE EGVVIRSENFDTNAKTIIVQLKESVEINCTRPN 271
Db 241 NVSTVQCTHGIRPVVSTQLLLNGSLAE EGVVIRSDNFTNAKTIIVQLKEPVEINCTRPN 300

Qy 272 NNTRKSIITGPGRAFATGDIIGDIRQAH CNISGEKWNNTLKQIVTKLQAQFGNKTIIVFK 331
Db 301 NNTRKSIHIGPGRAFYTGTGDIIGDIRQAH CNLSITKWNNTLKQIVKKLKEQFGNKTIIVFK 360

Qy 332 QSSGDPPEIVMHSFNCGGEFFCYNSTQLFNSTWNN-----TIGPNTNGTITLPCRICKQI 386
Db 361 QSSGDPPEIVTHSFNCGGEFFCYNSTKL FNSTWNNSTWNGTEGSNNPEGNITLPCRICKQI 420

Qy 387 INRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNT--TEIFRPGGDMRDNR 444
Db 421 VNMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGG---SNTSGTEIFRPGGDMRDNR 477

Qy 445 SELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQ 504
Db 478 SELYKYKVVKIEPLGVAPTAKRRVVQREKRAVIGIGALFLGFLGAAGSTMGAASMTLTVQ 537

Qy 505 ARQLLSGIVQQNNLLRAIEAQOHLQLTVWG IK-LQARVLAVERYLKDQQLLGWCSG 563
Db 538 ARLLLSGIVQQNNLLRAIEAQOHLQLTVWG IKQLQARVLAVERYLKDQQLLGWCSG 597

Qy 564 KLICTTAVPWNASWSNKS LDIWNNMTWMEWEREIDNYTNLIYTLIEESQNQQEKNEQEL 623
Db 598 KLICTTVPWNASWSNKS LDKIWDNMTWMEWEREINNYTSLIYTLIEESQNQQEKNELEL 657

Qy 624 LECLKWASLWNWFDISKWLWYIK 646
Db 658 LECLKWASLWNWFDTIKWLWYIK 680

RESULT 7
O40222
ID O40222 PRELIMINARY; PRT; 854 AA.
AC O40222;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Env polyprotein.

Db 361 SGGDEIVTHSFNCGGEFFCYNSTQLFNSTWNTDTEGSNDTERTITLPCRICKQIINMWQKV 420
Qy 394 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKV 453
Db 421 GKAMYAPPPIRGQIRCSSNITGLLLTRDGGNNNNNGTEIFRPVGGDMRDNRSELYKYKV 480
Qy 454 KIEPLGVAPTAKARRVVQREKRAVTLGAMFLGLGAAGSTWGARSLLTLTVQARQLLSGIV 513
Db 481 KIEPLGVAPTAKARRVVQREKRAVGLGAMFLGLGAAGSTWGARRLLTLTVQARQLLSGIV 540
Qy 514 QQQNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLGIWCGSKLICITTA VP 572
Db 541 QQQNLLRAIEAQHLLQLTVWGIKLQARVLAVERYLKDQILGIWCGSKSICTTAVP 600
Qy 573 WNASWSNKS LDIWNNMTWMEWEREIDNYTNLIYTLIEESQKQEKNEQELLELDKWASL 632
Db 601 WNASWSNKS LDIWNNMTWMEWEREIDNYTNLIYTLIEESQKQEKNEQELLELDKWASL 659
Qy 633 WNWFDISKWLWYIK 646
Db 660 WNWFDITRWLWYIK 673
RESULT 9
Q75760 PRELIMINARY; PRT; 847 AA.
AC Q75760;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JRFL;
RX MEDLINE=87206194; PubMed=3646751;
RA Koyanagi Y., Miles S., Mitsuyasu R.T., Merrill J.E., Vinters H.V.,
RA Chen I.S.;
RT "Dual infection of the central nervous system by AIDS viruses with
RT distinct cellular tropisms.";
RL Science 236:819-822(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JRFL;
RX MEDLINE=91043044; PubMed=2172833;
RA O'Brien W.A., Koyanagi Y., Namazie A., Zhao J.Q., Diagne A., Idler K.,
RA Zack J.A., Chen I.S.;
RT "HIV-1 tropism for mononuclear phagocytes can be determined by regions
RT of gp120 outside the CD4-binding domain.";
RL Nature 348:69-73(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JRFL;
RX MEDLINE=92092169; PubMed=1684385;
RA Pang S., Vinters H.V., Akashi T., O'Brien W.A., Chen I.S.;
RA "HIV-1 env sequence variation in brain tissue of patients with AIDS-
RT related neurologic disease.";
RL J. Acquir. Immune Defic. Syndr. 4:1082-1092(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=JRFL;
RA Pang S., Vinters H.V., Akashi T., O'Brien W.A., Chen I.S.,
RA Koyanagi Y., Namazie A., Zhao J., Diagne A., Idler K.,
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U63632; AAB05604.1; -.
DR PIR; S13289; S13289.
DR PIR; T09448; T09448.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 847 AA; 96160 MW; 022D5F24E04FB29F CRC64;
Query Match 87.4%; Score 3035; DB 15; Length 847;
Best Local Similarity 85.4%; Pred. No. 3.7e-240;
Matches 577; Conservative 29; Mismatches 38; Indels 32; Gaps 6;
Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLMICS AVEKLVWTVYGVVWKEATTLFCASDAKAY 60
Db 1 MRVKGIRKSYQYLWKGGTLLGLMICS AVEKLVWTVYGVVWKEATTLFCASDAKAY 60
Qy 61 DTEVHNVWATHACVPTDPNPQEI VLENVTENFNMMKNNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNVWATHACVPTDPNPQEVLENVTEHFNMMKNNMVEQMEDIISLWDQSLKPCVK 120
Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSPKVGAG----- 162
Db 121 LTPLCVTLNCKDV-NATNTNDSEGTMERGEIKNCSEFNITTSIRDEVQKEYALFYKLDVV 179
Qy 163 -----KLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNV 213
Db 180 PIDNNNTSYRLISCDTSVITQACPKISFEPIPIHYCAPAGFAILKCNCKTENGKGPCKNV 239
Qy 214 STVQCTHGIRPVVSTQLLNGSLAEVGVIRSENFTDNAKTIIVQLKESVEINCTRPNNN 273
Db 240 STVQCTHGIRPVVSTQLLNGSLAEVGVIRSDNFTNNAKTIIVQLKESVEINCTRPNNN 299
Qy 274 TRKSITIGPGRAFYATGDIIGDIRQAHNCNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQS 333
Db 300 TRKSIHIGPGRAFYTTGEIIGDIRQAHNCNISRAKWNNDTLKQIVIKLREQFENKTIIVFNHS 359
Qy 334 SGGDPEIVMHSFNCGGEFFCYNSTQLFNSTW-NNTIGPNNNG-TITLPCRICKQIINRWQ 391
Db 360 SGGDPEIVMHSFNCGGEFFCYNSTQLFNSTWNNNTEGNNTEGNTITLPCRICKQIINMWQ 419
Qy 392 EVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYK 451
Db 420 EVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGIN-ENGTEIFRPGGDMRDNRSELYKYK 478
Qy 452 VVKIEPLGVAPTAKARRVVQREKRAVTLGAMFLGLGAAGSTWGARSLLTLTVQARQLLSG 511
Db 479 VVKIEPLGVAPTAKARRVVQREKRAVGIGAVFLGLGAAGSTWGASMTLTVQARLLLSG 538
Qy 512 IVQQNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLGIWCGSKLICITTA 570
Db 539 IVQQNNLLRAIEAQHLLQLTVWGIKLQARVLAVERYLGDQQLGIWCGSKLICITTA 598
Qy 571 VPWNASWSNKS LDIWNNMTWMEWEREIDNYTNLIYTLIEESQKQEKNEQELLELDKWA 630
Db 599 VPWNASWSNKS LDIWNNMTWMEWEREIDNYTSEIYTLIEESQKQEKNEQELLELDKWA 658
Qy 631 SLWNWFDISKWLWYIK 646
Db 659 SLWNWFDITRWLWYIK 674
RESULT 10
Q70150 PRELIMINARY; PRT; 843 AA.
ID Q70150
AC Q70150;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.


```
OS Human immunodeficiency virus 1.
OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SFMS3.1;
RX MEDLINE=98178716; PubMed=9519894;
RA McCutchan F.E., Sanders-Buell E., Salminen M.O., Carr J.K.,
RA Sheppard W.H.;
RT "Diversity of the human immunodeficiency virus type 1 envelope
RT glycoprotein in San Francisco Men's Health Study participants.";
RL AIDS Res. Hum. Retroviruses 14:329-337(1998).
DR EMBL; AF025751; AAC40589.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR00328; Env_GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 853 AA; 96830 MW; EC8BA4418F8AEC47 CRC64;

Query Match      87.0%; Score 3022; DB 15; Length 853;
Best Local Similarity 84.4%; Pred. No. 4.4e-239;
Matches 574; Conservative 25; Mismatches 47; Indels 34; Gaps 4;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLICSAVEKLVVTYYGVVPVWKEATTLFCASDAKAY 60
Db 1 MRVKEIRKNYQHLWRGWTMLGLMLICSAAEKLVVTYYGVVPVWKEATTLFCASDAKAY 60

Qy 61 DTEVHNWVWATHACVPTDPNPQEIIVLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 DTEVHNWVWATHACVPTDPNPQEVGLENTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCT-NLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db 121 LTPLCVTLNRTDCLKNATNTNSSGEMEMRGEIKNCSFNITTSIRDKVKEYALFYKLDVV 180

Qy 163 -----KLINCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPC 210
Db 181 PINNNNTTTSYRLINCNTSVITQACPVSFEPIPIHYCTPAGFAILKCKDKKFNCGTGPC 240

Qy 211 TNVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFNTDIAKIIIVQLKESVEINCTRP 270
Db 241 TSVSTVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFNTDIAKIIIVQLNESVESNCTRP 300

Qy 271 NNTRKSIITIGPRAFYATGDIIGDIRQAHNCISGEKWNNTLKOIVTKLQAQFGNKTIIV 330
Db 301 NNTRKSIPIGPRAFYTTGEIIGDIRQAHNCISRTKWDNALQIAKLGKQFGTKTIIV 360

Qy 331 KQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSNTWNTIGPNT---NGTITLPCRICKII 387
Db 361 NQSSGGDPEIVMHTFNCGGEFFYCNSTQLFNSNTWNTRESNNSSTEGNTITLPCRICKII 420

Qy 388 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDMRDNRSEL 447
Db 421 NMWQGVKAMYAPPPIRGQIRCSSNITGLLLTRDGGNNESNETETFRPGGGDMRDNRSEL 480

Qy 448 YKYKVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 507
Db 481 YKYKVKIEPLGVAPTAKRRVVQREKRAVGIGAVFLGFLGAAGSTMGAASMTLTVQARL 540

Qy 508 LLSGIVQOQNLLRAIEAQOHLQLTVWGIK-LQARVLAVERYLKDQQLGIWCGSGKLI 566
Db 541 LLSGIVQOQNLLRAIEAQOHLQLTVWGIKQLQARVLAVERYLKDQQLGIWCGSGKLI 600

Qy 567 CTTAVPWNASWSNKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLEL 626
Db 601 CTTVPWNASWSNKSLDKIWHNTWMEWEREIDNYTSLIYTLIEESNQOEKNELELLEL 660

Qy 627 DKWASLWNWFDISKWLWYIK 646
```

```
Db 661 DKWASLWNWFDITKWLWYIK 680

RESULT 14
O56110 PRELIMINARY; PRT; 851 AA.
AC O56110;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SFMS8.1;
RX MEDLINE=98178716; PubMed=9519894;
RA McCutchan F.E., Sanders-Buell E., Salminen M.O., Carr J.K.,
RA Sheppard W.H.;
RT "Diversity of the human immunodeficiency virus type 1 envelope
RT glycoprotein in San Francisco Men's Health Study participants.";
RL AIDS Res. Hum. Retroviruses 14:329-337(1998).
DR EMBL; AF025756; AAC40593.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. .; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR00328; Env_GP41.
DR InterPro; IPR00173; GAP_dhdrogenase.
DR InterPro; IPR00777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
DR PROSITE; PS00071; GAPDH; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 851 AA; 96516 MW; 1CDAC83EE6464531 CRC64;

Query Match      86.9%; Score 3020; DB 15; Length 851;
Best Local Similarity 83.2%; Pred. No. 6.4e-239;
Matches 565; Conservative 38; Mismatches 42; Indels 34; Gaps 4;

Qy 1 MRVKGIRKNYQHLWRGGTLLGLMLICSAVEKLVVTYYGVVPVWKEATTLFCASDAKAY 60
Db 1 MRVKEIRKNYQHLWRGWTMLGLMLICSAADKLVVTYYGVVPVWKEATTLFCASDAKAY 60

Qy 61 DTEVHNWVWATHACVPTDPNPQEIIVLENVTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120
Db 61 ETEVHNWVWATHACVPTDPDPQEVLENTENFNMWKNMVEQMHEDIISLWDQSLKPCVK 120

Qy 121 LTPLCVTLHCTNLKNATNTKSSNWKEMDRGEIKNCSFKVGAG----- 162
Db 121 LTPLCVTLNCIDLNDTNTSSGETMERGEIKNCSFNITTSIRDKVKEYALLHKLDVV 180

Qy 163 -----KLINCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPCTNVS 214
Db 181 PIDNTSYRLVSCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNCKRFGKGPCTNVS 240

Qy 215 TVQCTHGIRPVVSTQLLNGSLAEEGVVIRSENFNTDIAKIIIVQLKESVEINCTRPNNNT 274
Db 241 TVQCTHGIRPVVSTQPLNGSLAEEKVIRSDNTDIAKIIIVQLNESVEINCTRPNNNT 300

Qy 275 RKSITIGPRAFYATGDIIGDIRQAHNCISGEKWNNTLKOIVTKLQAQFGNKTIIVFKQSS 334
Db 301 RKSINIGPRAFYTTGEIIGDIRQAHNCNLSEAKWNHTLEQIAKKLREQFGNKTIIVFNQSS 360

Qy 335 GGDPEIVMHSFNCGGEFFYCNSTQLF-----NSTWNTTIGPNNNTGTITLPCRICKIIN 388
Db 361 GGDPEIVMYSFNCGGEFFYCNSTKLFNRTWSVNSTWNTDEGVNNTGENITLPCRICKIIN 420
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: February 25, 2004, 14:02:44 ; Search time 52.8474 Seconds
(without alignments)
3309.468 Million cell updates/sec

Title: US-09-891-609A-4
Perfect score: 3327
Sequence: 1 SAVEKLWVTYYGVVPVWKEA.....DKWASLWNWFDISKWLWYIK 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3327	100.0	619	5 AAU75156	Aau75156 N-termina
2	3327	100.0	646	5 AAU75155	Aau75155 Modified
3	3275	98.4	842	6 ABR55684	Abr55684 HIV isola
4	3275	98.4	842	7 ADC13218	Adcl3218 Protein o
5	3275	98.4	847	3 AAY97073	Aay97073 Variant H
6	3270	98.3	842	6 ABU66565	Abu66565 Human imm
7	3260	98.0	842	5 ABB06211	Abb06211 HIV Env i
8	2942.5	88.4	850	2 AAR67724	Aar67724 gpl20 fro
9	2928	88.0	883	4 AAB82761	Aab82761 Ancestral
10	2919	87.7	856	6 ABR55495	Abr55495 Amino aci
11	2908.5	87.4	851	1 AAP80967	Aap80967 HIV prote
12	2876.5	86.5	855	2 AAW11581	Aaw11581 Human Imm
13	2876.5	86.5	855	2 AAW88113	Aaw88113 Env prote
14	2862	86.0	643	4 AAB61505	Aab61505 HIV-1 SOS
15	2862	86.0	643	6 ABR57052	Abr57052 HIV-1 JR-
16	2856	85.8	855	2 AAW53112	Aaw53112 ENV prote
17	2856	85.8	855	3 AAY77298	Aay77298 HIV-1 (AT
18	2856	85.8	855	3 AAY77302	Aay77302 HIV-1 (AT
19	2856	85.8	855	6 ABU57553	Abu57553 AIDS asso
20	2856	85.8	855	6 ABU57550	Abu57550 AIDS asso
21	2856	85.8	855	6 ABU63182	Abu63182 Protein #
22	2856	85.8	855	6 ABU63186	Abu63186 Env prote
23	2856	85.8	863	1 AAP61509	Aap61509 Sequence
24	2856	85.8	863	2 AAR29706	Aar29706 env gene
25	2856	85.8	863	5 AAE35790	Aae35790 ARV-2 (9B

26	2851	85.7	856	4 AAB85999	Aab85999 Amino aci
27	2851	85.7	863	2 AAR43869	Aar43869 HTLV-III
28	2848.5	85.6	868	5 AAO19389	Aao19389 Lymphaden
29	2845.5	85.5	868	1 AAP60422	Aap60422 Sequence
30	2844	85.5	857	2 AAR67725	Aar67725 gpl20 fro
31	2843	85.5	856	2 AAR25940	Aar25940 Modified
32	2843	85.5	866	1 AAP80966	Aap80966 HIV prote
33	2842.5	85.4	880	2 AAW23333	Aaw23333 Human imm
34	2842.5	85.4	880	2 AAW73332	Aaw73332 HIV envel
35	2842.5	85.4	880	3 AAB14844	Aab14844 HIV envel
36	2841.5	85.4	868	1 AAP60063	Aap60063 HIV virus
37	2841	85.4	863	1 AAP60349	Aap60349 HTLV-III
38	2838	85.3	856	1 AAP61514	Aap61514 Sequence
39	2838	85.3	856	2 AAW89325	Aaw89325 HIV-1 env
40	2838	85.3	856	6 ABU63322	Abu63322 Human lym
41	2835	85.2	854	3 AAB10053	Aab10053 HIV-1 coa
42	2835	85.2	854	3 AAB10697	Aab10697 HIV-1 env
43	2835	85.2	854	4 AAB86199	Aab86199 HIV gp41
44	2835	85.2	854	5 ABG73663	Abg73663 HIV-1 NL4
45	2834	85.2	856	3 AAY97072	Aay97072 Wild type

ALIGNMENTS

RESULT 1
AAU75156
ID AAU75156 standard; protein; 619 AA.
XX

AC AAU75156;

XX
DT 08-MAY-2002 (first entry)

XX
DE N-terminal mutant of modified HIV-1 SF162DV2 gpl40 envelope protein.

XX
KW Human immunodeficiency virus type 1; HIV-1; gpl40 envelope protein;

XX
KW second hypervariable region; V2; SF162deltav2; SF162DV2; AIDS;

XX
KW acquired immunodeficiency syndrome; anti-HIV; mutant; mutein; immunogen.

OS Human immunodeficiency virus 1; strain SF162 (clade B).

OS Synthetic.

XX
FH Key Location/Qualifiers

FT Misc-difference 512

FT /note= "Encoded by CAGCTC"

XX
PN WO200200250-A2.

XX
PD 03-JAN-2002.

XX
PF 27-JUN-2001; 2001WO-US020483.

XX
PR 27-JUN-2000; 2000US-0214608P.

XX
PR 26-JUN-2001; 2001US-00891609.

XX
PA (AARO-) AARON DIAMOND AIDS RES CENT.

XX
PI Stamatos L, Barnett S, Shrivastava I;

XX
DR WPI; 2002-130836/17.

XX
DR N-PSDB; AAS15499.

XX
PT Immunizing an animal and eliciting an immune response against

PT heterologous HIV-1 in an animal, involves administering an immunogen

PT comprising modified HIV-1 envelope protein, or DNA or virus encoding the

PT protein.

XX
PS Claim 5; Fig 17; 62pp; English.

XX
CC The present invention relates to methods for immunising an animal,

CC preferably human, against heterologous human immunodeficiency virus type

CC 1 (HIV-1), and eliciting a heterologous immune response to HIV-1 in that

CC animal. The method comprises administering an immunogen having at least

Db 268 TRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAHNCNISGEKWNNTLKQIVTKLQAQFGNKT 327
Qy 301 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSWTNNNTIGPNNNTNGTITLPCRICKQII 360
Db 328 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSWTNNNTIGPNNNTNGTITLPCRICKQII 387
Qy 361 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDMRDNRSEL 420
Db 388 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDMRDNRSEL 447
Qy 421 YKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 480
Db 448 YKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 507
Qy 481 LLSGIVQQNNLLRAIEAQHLLQLTVWGIKLQARVLAVERYLKDQQLLGWCSGKLIC 540
Db 508 LLSGIVQQNNLLRAIEAQHLLQLTVWGIKLQARVLAVERYLKDQQLLGWCSGKLIC 567
Qy 541 TTAVPWNASWSKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQEKNEQELLELD 600
Db 568 TTAVPWNASWSKSLDQIWNNTWMEWEREIDNYTNLIYTLIEESQNQEKNEQELLELD 627
Qy 601 KWASLWNWFDISKWLWYIK 619
Db 628 KWASLWNWFDISKWLWYIK 646
RESULT 3
ABR55684
ID ABR55684 standard; protein; 842 AA.
XX
AC ABR55684;
XX
DT 23-OCT-2003 (revised)
DT 18-AUG-2003 (first entry)
XX
DE HIV isolate SF162 env polypeptide.
KW HIV; gag; nef; prot; tat; rev; vif; vpr; vpu; env; anti-HIV; vaccine;
KW immune response.
XX
OS Human immunodeficiency virus 1.
XX
PN WO2003020876-A2.
XX
PD 13-MAR-2003.
XX
PF 05-JUL-2002; 2002WO-US021342.
XX
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-278761/27.
XX
PT New expression cassettes and polynucleotides encoding HIV Gag, Nef, Prot,
PT Tat, Rev, Vif, Vpr, Vpu, or Env polypeptides, useful for DNA immunization
PT or generating an immune response against HIV in a subject.
XX
PS Example; Fig 2A-C; 214pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Nef,
CC Prot, Tat, Rev, Vif, Vpr, Vpu, or Env polypeptide. The expression
CC cassettes, HIV polypeptides and polynucleotides encoding the HIV
CC polypeptides are useful for DNA immunization or generating an immune
CC response against HIV in a subject. The polynucleotides are also useful
CC for generating packaging cell lines or producing the HIV polypeptides.
CC Sequences ABR55684-688 represent env polypeptides from various HIV

CC isolates. (Updated on 23-OCT-2003 to standardise OS field)
XX
SQ Sequence 842 AA;
Query Match 98.4%; Score 3275; DB 6; Length 842;
Best Local Similarity 95.2%; Pred. No. 5e-159;
Matches 616; Conservative 0; Mismatches 3; Indels 28; Gaps 2;
Qy 1 SAVEKLLWVTVYYGVVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60
Db 23 SAVEKLLWVTVYYGVVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 82
Qy 61 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
Db 83 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 142
Qy 121 DRGEIKNCSFKVGAG-----KLNCSNTSVITQACPKVS 153
Db 143 DRGEIKNCSFKVTTTSIRNMKQKEYALFYKLDVVPIDNDNTSVYKLNCSNTSVITQACPKVS 202
Qy 154 FEPIPIHYCAPAGFAILKCNCKKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 213
Db 203 FEPIPIHYCAPAGFAILKCNCKKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 262
Qy 214 VVIRSENFTDNAKTIIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273
Db 263 VVIRSENFTDNAKTIIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 322
Qy 274 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 333
Db 323 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 382
Qy 334 NSTWNNTIGPNNNTNGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 393
Db 383 NSTWNNTIGPNNNTNGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 442
Qy 394 GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLG 453
Db 443 GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLG 502
Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
Db 503 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQL 562
Qy 513 QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSKSLDQIWNNTWMEWEREID 572
Db 563 QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSKSLDQIWNNTWMEWEREID 622
Qy 573 NYTNLIYTLIEESQNQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 623 NYTNLIYTLIEESQNQEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 4
ADC13218
ID ADC13218 standard; protein; 842 AA.
XX
AC ADC13218;
XX
DT 18-DEC-2003 (first entry)
XX
DE Protein of HIV Type C, SF162 SEQ ID NO 2.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX WPI; 2003-221593/21.
DR
XX New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 2; 301pp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramusosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This sequence represents an HIV Type C related protein of
CC the invention.
XX
SQ Sequence 842 AA;

Query Match 98.4%; Score 3275; DB 7; Length 842;
Best Local Similarity 95.2%; Pred. No. 5e-159;
Matches 616; Conservative 0; Mismatches 3; Indels 28; Gaps 2;

Qy 1 SAVEKLMWTVYYGVVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60
Db
23 SAVEKLMWTVYYGVVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 82

Qy 61 VTENFNNWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
Db
83 VTENFNNWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 142

Qy 121 DRGEIKNCSFKVGAG-----KLINCNTSVITOACPKVS 153
Db
143 DRGEIKNCSFKVTTSIRNKMQKEYALFYKLDVVPIDNDNTSYKLINCNTSVITOACPKVS 202

Qy 154 FEPIPIHYCAPAGFAILKCNDDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEG 213
Db
203 FEPIPIHYCAPAGFAILKCNDDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEG 262

Qy 214 VVIRSENFDTNAKTIIVQLKESVEINCTRPNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273
Db
263 VVIRSENFDTNAKTIIVQLKESVEINCTRPNNTRKSIITIGPGRAFYATGDIIGDIRQAH 322

Qy 274 CNISGEKNWNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 333
Db
323 CNISGEKNWNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 382

Qy 334 NSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGIRCSSNITGLLLTRD 393
Db
383 NSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGIRCSSNITGLLLTRD 442

Qy 394 GGKEISNTTEIFRPGGDMRDNRSELYKYVVKIEPLGVAPTAKRRVWQREKRAVTLG 453
Db
443 GGKEISNTTEIFRPGGDMRDNRSELYKYVVKIEPLGVAPTAKRRVWQREKRAVTLG 502

Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGK-L 512
Db
503 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGKQL 562

Qy 513 QARVLAVERYLKDQQLGIWGCSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREID 572
Db
563 QARVLAVERYLKDQQLGIWGCSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREID 622

Qy 573 NYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db
623 NYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 5
AAY97073
ID AAY97073 standard; protein; 847 AA.
XX
AC AAY97073;
XX
DT 12-SEP-2003 (revised)
DT 31-OCT-2000 (first entry)
XX
DE Variant HIV-1 SF162 Env gpl60.
XX
KW HIV-1; vaccine; anti-HIV; immunogenic; modified; envelope; Env;
KW CD4 binding region; V1/V2 loop; bridging sheet.
XX
OS Human immunodeficiency virus 1; isolate SF162.
XX
PN WO200039303-A2.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US031272.
XX
PR 31-DEC-1998; 98US-0114495P.
PR 29-SEP-1999; 99US-0156670P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Barnett S, Hartog K, Martin E;
XX
DR WPI; 2000-465745/40.
XX
PT Novel modified HIV Env polypeptides useful as immunizing agents and for
PT preparing a vaccine to elicit an immune response against a broad range of
PT HIV subtypes.
XX
PS Claim 5; Page 115-117; 139pp; English.
XX
CC Novel immunogenic modified human immunodeficiency virus (HIV) envelope
CC (Env) polypeptides having an amino acid deleted or replaced in the region
CC corresponding to residues 420-436 or 119-123 and 199-210 relative to
CC isolate HXB-2 are disclosed. The modified Env polypeptide is based on HIV
CC strain SF162, with numbering relative to isolate HXB-2. The Env
CC polypeptides are modified so as to expose at least part of the CD4
CC binding region. The modified HIV Env polypeptides, coding polynucleotides
CC and constructs, further comprising an adjuvant, are used for inducing an
CC immune response in an individual. The method involves administering a
CC first composition comprising a polynucleotide encoding the Env
CC polypeptide in a priming step and administering a second composition
CC comprising a modified Env polypeptide as a booster in an amount
CC sufficient to induce an immune response in the individual. The first
CC and/or second composition further comprises an adjuvant (claimed). The
CC intracellularly produced Env polypeptides can be used for a number of
CC diagnostic and therapeutic purposes to determine the presence of reactive
CC antibodies and/or Env proteins in a biological sample to aid in the
CC diagnosis of HIV infection or disease status or as measure of response to
CC immunization. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 847 AA;

Query Match 98.4%; Score 3275; DB 3; Length 847;
Best Local Similarity 95.2%; Pred. No. 5e-159;
Matches 616; Conservative 0; Mismatches 3; Indels 28; Gaps 2;

Qy 1 SAVEKLMWTVYYGVVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60

Db 28 SAVEKLWVTYYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 87
Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
Db 88 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 147
Qy 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACPVS 153
Db 148 DRGEIKNCSFKVTTTSIRNKMQKEYALFYKLDVVPIDNDNTSYKLINCNTSVITQACPVS 207
Qy 154 FEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAG 213
Db 208 FEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAG 267
Qy 214 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273
Db 268 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNTRKSIITIGPGRAFYATGDIIGDIRQAH 327
Qy 274 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 333
Db 328 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 387
Qy 334 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 393
Db 388 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 447
Qy 394 GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLG 453
Db 448 GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLG 507
Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
Db 508 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQL 567
Qy 513 QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSILDIWNNMTWMEWEREID 572
Db 568 QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSILDIWNNMTWMEWEREID 627
Qy 573 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNNWFDISKWLWYIK 619
Db 628 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNNWFDISKWLWYIK 674

RESULT 6
ABU66565
ID ABU66565 standard; protein; 842 AA.

AC ABU66565;
XX
DT 22-MAY-2003 (first entry)
XX
DE Human immunodeficiency virus (HIV) envelope (env) protein #1.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation;
KW envelope protein; env.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX

PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
XX Disclosure; Fig 2A-C; 262pp; English.
XX
XX The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This is the amino acid
CC sequence of a human immunodeficiency virus (HIV) envelope (env) protein
XX
SQ Sequence 842 AA;

Query Match 98.3%; Score 3270; DB 6; Length 842;
Best Local Similarity 95.1%; Pred. No. 9e-159;
Matches 615; Conservative 0; Mismatches 4; Indels 28; Gaps 2;
Qy 1 SAVEKLWVTYYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60
Db 23 SAVEKLWVTYYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 82
Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
Db 83 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 142
Qy 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACPVS 153
Db 143 DRGEIKNCSFKVTTTSIRNKMQKEYALFYKLDVVPIDNDNTSYKLINCNTSVITQACPVS 202
Qy 154 FEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAG 213
Db 203 FEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAG 262
Qy 214 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273
Db 263 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNTRKSIITIGPGRAFYATGDIIGDIRQAH 322
Qy 274 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 333
Db 323 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 382
Qy 334 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 393
Db 383 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 442
Qy 394 GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLG 453
Db 443 GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKTEPLGVAPTKAKRRVVQREKRAVTLG 502
Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
Db 503 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQL 562
Qy 513 QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSILDIWNNMTWMEWEREID 572
Db 563 QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSILDIWNNMTWMEWEREID 622
Qy 573 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNNWFDISKWLWYIK 619

RESULT 14					
ID	AAB61505	standard; protein; 643 AA.			
XX	AC	AAB61505;			
XX	DT	11-SEP-2003 (revised)			
DT	05-APR-2001	(first entry)			
XX	DE	HIV-1 SOS gp140 glycoprotein.			
XX	KW	gp120; anti-HIV; gene therapy; vaccine; viral envelope protein; gp41; gp140.			
XX	OS	Human immunodeficiency virus 1.			
PN	WO200100648-A1.				
XX	PD	04-JAN-2001.			
XX	Pf	23-JUN-2000; 2000WO-US017267.			
XX	PR	25-JUN-1999; 99US-00340992.			
XX	PA	(PROG-) PROGENICS PHARM INC.			
XX	PA	(AARO-) AARON DIAMOND AIDS RES CENT.			
XX	PI	Binley JM, Schuelke N, Olson WC, Maddon PJ, Moore JP;			
XX	DR	WPI; 2001-122993/13.			
XX	DR	N-PSDB; AAF28581.			
XX	PT	New viral envelope proteins, useful for producing vaccines to treat human immunodeficiency virus-1 infections, comprises amino acid sequence mutations such that viral transmembrane-surface protein complex is more stable.			
PS	Fig 13;	109pp; English.			
CC	The present invention relates to a viral envelope protein. The viral envelope protein comprises a viral surface protein (e.g. glycoprotein gp120) and a corresponding viral transmembrane protein (e.g. gp41), in which the viral envelope protein contains one or more amino acid sequence mutations that enhance the stability of the complex formed between the viral surface and transmembrane proteins. The viral envelope protein can be used in the treatment of viral infection e.g. HIV-1 infection. The present sequence is HIV-1 SOS gp140, which was used in the present invention. (Updated on 11-SEP-2003 to standardise OS field)				
SQ	Sequence	643 AA;			
Query Match 86.0%; Score 2862; DB 4; Length 643;					
Best Local Similarity 84.5%; Pred. No. 4.2e-138;					
Matches 545; Conservative 25; Mismatches 43; Indels 32; Gaps 6					
Qy	3	VEKLWTVVYGGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIVLENTV 62			
Dd	1	VEKLWTVVYGGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIVLENTV 60			
Qy	63	ENFNMWKNMVEQMEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEMDR 122			
Dd	61	EHFNMWKNMVEQMEDIISLWDQSLKPCVKLTPLCVTLNCCKDV-NATNTTNDSEGTMER 119			
Qy	123	GEIKNCSPKVAG-----KLINCNTSVITQACPVSFE 155			
Dd	120	GEIKNCSEFNIITSIRDEVQKEYALFYKLDVVPIDNNNTSVRLISCDTSVITQACPKISFE 179			
Qy	156	PIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCETHGIRPVVSTQLLLNGSLAEGLV 215			
Dd	180	PIPIHYCAPAGFAILKCNDKTENGKGPCKNVSTVQCETHGIRPVVSTQLLLNGSLAEGLV 239			

QY	216	IRSEFTDNAKTIIVQLKESVEINCTRPNNNTRKSITIGPGRAFYATGDIIGDIRQAHCN	278
		: : : : : : : : :	
Dd	240	IRSDNFTNNAKTIIVQLKESVEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHCN	299
		: : : : : : : : :	
QY	276	ISGEKWNTLKQIVTKLQAQFGNKTIIVFKOSSGGDPEIVMHSFNCGGEFFYCNSQTFLFNS	335
		: : : : : : : : :	
Dd	300	ISRAKWNDTLKQIVIKLREQFENKTIVFNHSSGGDPEIVMHSFNCGGEFFYCNSQTFLFNS	359
		: : : : : : : : :	
QY	336	TW-NNTIGPNNTNG-TITLPCRIKQIIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLLTRD	393
		: : : : : : : : :	
Dd	360	TWNNTTEGSNNTGNITITLPCRIKQIIINMWQEVGKAMYAPPPIRGQIRCSSNITGLLLLTRD	419
		: : : : : : : : :	
QY	394	GGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTKAKRRVVQREKRAVTLG	453
		: : : : : : : : :	
Dd	420	GGIN-ENGTEIFRPGGDMRDNRSEFYKYKVVKIEPLGVAPTKCKRRVVQREKRAVGIG	478
		: : : : : : : : :	
QY	454	AMFLGFLGAAGSTMARSLTITVQARQLLSGIVQQQNLLRAIEAQOHLQLTVWGIL-L	512
		: : : : : : : : :	
Dd	479	AVFLGFLGAAGSTMGAASMTITVQARLLLSGIVQQQNLLRAIEAQORMLQLTVWGIKQL	538
		: : : : : : : : :	
QY	513	QARVLAVERYLKDQQLGIWCSGKLICTTAVPWNASWSNKSLLDQIWNNMTWMWEEREID	572
		: : : : : : : : :	
Dd	539	QARVLAVERYLGDQQLGIWCSGKLICCTAVPWNASWSNKSLLDRINNNMTWMWEEREID	598
		: : : : : : : : :	
QY	573	NYTNLIYTLIESQNOQEKNELLEDKWLWNLWFDISKWLWY	617
		: : : : : : : : :	
Dd	599	NYTSEIYTLIESQNOQEKNELLEDKWLWNLWFDTNWNLWY	643
		: : : : : : : : :	
 RESULT 15 ABR57052			
ID	ABR57052	standard; protein; 643 AA.	
XX			
AC	ABR57052;		
XX			
DT	23-OCT-2003	(revised)	
DT	05-AUG-2003	(first entry)	
XX			
DE	HIV-1 JR-FL SOS gp140 amino acid sequence.		
XX			
KW	Human immunodeficiency virus; envelope glycoprotein trimeric complex;		
KW	HIV; anti-HIV; vaccine; immune response; HIV infection; gp120; gp41;		
KW	gp140; furin-recognition sequence.		
XX			
OS	Human immunodeficiency virus 1.		
XX			
PN	WO2003022869-A2.		
PD	20-MAR-2003.		
XX			
PF	06-SEP-2002; 2002WO-US028331.		
XX			
PR	06-SEP-2001; 2001US-0317764P.		
PR	06-SEP-2001; 2001US-0317775P.		
PR	06-SEP-2001; 2001US-0317909P.		
PR	06-SEP-2001; 2001US-0317910P.		
PR	05-APR-2002; 2002US-0370264P.		
PR	05-APR-2002; 2002US-0370410P.		
XX			
PA	(PROG-) PROGENICS PHARM INC.		
PA	(CORR) CORNELL RES FOUND INC.		
XX			
PI	Moore JP, Binley JM, Lu M, Olson WC, Schulke N, Gardner J;		
PI	Maddon PJ, Sanders R;		
XX			
DR	WPI; 2003-371744/35.		
DR	N-PSDB; ACC79597.		
XX			
PT	Novel stable HIV-1 pre-fusion envelope glycoprotein trimeric complex in		
PT	which each monomeric unit of the complex comprises HIV-1 gp120 and HIV-1		
PT	gp41, useful for eliciting immune response in subject against HIV-1.		
XX			
PS	Example; Fig 13B; 316pp; English.		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: February 25, 2004, 14:14:30 ; Search time 18.1051 Seconds
(without alignments)
1765.050 Million cell updates/sec

Title: US-09-891-609A-4
Perfect score: 3327
Sequence: 1 SAVEKLWTVVYGVVPVWKEA.....DKWASLWNWFDISKWLWYIK 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2942.5	88.4	850	2	US-08-448-603A-28
2	2942.5	88.4	850	3	US-09-134-075-28
3	2942.5	88.4	850	4	US-09-492-739-28
4	2880.5	86.6	855	3	US-07-956-483-14
5	2876.5	86.5	855	1	US-08-022-835-6
6	2876.5	86.5	855	1	US-08-388-809-6
7	2876.5	86.5	855	2	US-08-647-714-6
8	2851	85.7	856	3	US-09-124-900-9
9	2851	85.7	863	3	US-08-463-210-11
10	2851	85.7	863	4	US-08-463-028-11
11	2848.5	85.6	861	1	US-08-127-499A-14
12	2848.5	85.6	861	1	US-08-482-847-14
13	2848.5	85.6	861	3	US-07-956-483-10
14	2848.5	85.6	861	3	US-08-472-240A-1
15	2848.5	85.6	861	3	US-08-472-240A-7
16	2848.5	85.6	861	4	US-08-817-441-103
17	2848	85.6	855	3	US-07-956-483-15
18	2844	85.5	857	2	US-08-448-603A-30
19	2844	85.5	857	3	US-09-134-075-30
20	2844	85.5	857	4	US-09-492-739-30
21	2842.5	85.4	880	2	US-08-788-815-7
22	2842.5	85.4	880	3	US-09-157-963-7
23	2840.5	85.4	861	3	US-07-956-483-16
24	2839	85.3	865	3	US-07-956-483-13
25	2835	85.2	826	1	US-08-375-510-2
26	2835	85.2	826	2	US-08-487-657-2
27	2835	85.2	854	4	US-09-309-572-23

28	2835	85.2	854	4	US-09-718-096-23	Sequence 23, Appl
29	2834	85.2	856	2	US-07-916-098A-2	Sequence 2, Appli
30	2826	84.9	887	3	US-08-472-240A-5	Sequence 5, Appli
31	2819.5	84.7	839	3	US-08-472-240A-10	Sequence 10, Appl
32	2819	84.7	856	4	US-09-337-387-11	Sequence 11, Appl
33	2810.5	84.5	687	4	US-09-536-977-70	Sequence 70, Appl
34	2810.5	84.5	820	4	US-09-536-977-72	Sequence 72, Appl
35	2797.5	84.1	726	4	US-09-337-387-3	Sequence 3, Appli
36	2793.5	84.0	759	4	US-09-337-387-12	Sequence 12, Appl
37	2786	83.7	856	1	US-08-022-835-2	Sequence 2, Appli
38	2786	83.7	856	1	US-08-388-809-2	Sequence 2, Appli
39	2786	83.7	856	2	US-08-647-714-2	Sequence 2, Appli
40	2786	83.7	856	3	US-07-956-483-11	Sequence 11, Appl
41	2785	83.7	887	3	US-08-472-240A-6	Sequence 6, Appli
42	2782	83.6	857	1	US-08-022-835-4	Sequence 4, Appli
43	2782	83.6	857	1	US-08-388-809-4	Sequence 4, Appli
44	2782	83.6	857	2	US-08-647-714-4	Sequence 4, Appli
45	2782	83.6	887	3	US-08-472-240A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-448-603A-28
; Sequence 28, Application US/08448603A
; Patent No. 5864027
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; APPLICANT: Nakamura, Gerald R.
; TITLE OF INVENTION: HIV Envelope Polypeptides
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
; STREET: 3 Embarcadero Center
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,603A
; FILING DATE: 07-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/072,833
; FILING DATE: 07-JUN-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Haliday, Emily
; REGISTRATION NUMBER: 38903
; REFERENCE/DOCKET NUMBER: 14918-704
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-393-2000
; TELEFAX: 415-393-2286
; TELEX:
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-448-603A-28

Query Match 88.4%; Score 2942.5; DB 2; Length 850;
Best Local Similarity 85.3%; Pred. No. 8.1e-234;
Matches 557; Conservative 27; Mismatches 32; Indels 37; Gaps 5;

QY 1 SAVEKLWTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEI VLEN 60
||| |||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 151 KVSFEPPIHYCAPAGFAILKNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLA 210
Db 208 KVSFEPPIHYCA-RWFAILLNCNNKKFNGTGPCTNVSTVQCTHGIRPVVSTHLLNGSLA 266
Qy 211 EGVVIRSENFTDNAKTIIVQLKESVEINCTRPNNNRKSTITIGPGRAFYATGDIIGDIR 270
Db 267 EEEVLRSENFTDNAKTIIVQLKEAVEINCTRPNNNTTRSIIHGPGRAFYATGDIIGDIR 326
Qy 271 QAHCNISGEKWNNTLKQIVTKLOAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNST 330
Db 327 QAHCNISRAKWNNTLKQIVIKLRDPFNKTIIFNRSSGGDPEIVMHSFNCGGEFFYCNST 386
Qy 331 QLFNSTWNNTIGPNNNG--TITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGL 388
Db 387 QLFSSWTNGTEGSNTGNDTITLPCRICKQIINRWQEVGKAMYAPPPIKGQVKCSSNITGL 446
Qy 389 LLTRDGGKEISNT----TEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQ 444
Db 447 LLTRDGGNSKNGSKNENTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQ 506
Qy 445 REKRAV-TLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL 503
Db 507 REKRAVGITGAMFLGFLGAAGSTMGATSMTLTVQARLLSGIVQQNNLLRAIEAQOHL 566
Qy 504 QLTVMGK-LQARVLAVERYLKDQQLGIWCGSGKLICTTAVPWNASWSNKSOLDQIWN 562
Db 567 QLTVMGKQLQARVLAVERYLRDQQLGIWCGSGKLICTTVPWNTSWSNKSOLDKIWN 626
Qy 563 TWMEWERIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 627 TWMEWERIDNYTSLIYTLIEESQOQEKNEQELLELDKWASLWNWFNITNLWYIK 683

RESULT 5

US-08-022-835-6
; Sequence 6, Application US/08022835
; Patent No. 5420030
; GENERAL INFORMATION:
; APPLICANT: Reitz Jr., Marvin S.
; APPLICANT: Franchini, Genoveffa
; APPLICANT: Markham, Phillip D.
; APPLICANT: Gallo, Robert C.
; APPLICANT: Popovic, Mikulas
; APPLICANT: Garnter, Suzanne
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: Eleventh Floor, 1615 L. Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/022,835
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/599,491
; FILING DATE: 17-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Scott, Watson T.
; REGISTRATION NUMBER: 26,581
; REFERENCE/DOCKET NUMBER: WTS/5683/78507/SRL
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 855 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-022-835-6

Query Match 86.5%; Score 2876.5; DB 1; Length 855;
Best Local Similarity 82.7%; Pred. No. 2.2e-228;
Matches 542; Conservative 33; Mismatches 43; Indels 37; Gaps 5;

Qy 1 SAVEKLMVTVVYGVPMKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEI 60
Db 28 NAEKLMVTVVYGVPMKEATTTLFCASDRKAYDTEVHNWATHACVPTDPNPQEVELKN 87
Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTKSSNWKEM 120
Db 88 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNATNGNDTNTSS 147
Qy 121 DR-----GEIKNCSEFKVGAG-----KLINCTSVITQ 147
Db 148 SRGMVGGEMKNCSEFNITNIRGKVQKEYALFYKLDIAPIDNNSNNRYRLISCVITQ 207
Qy 148 ACPKVSFEPIPIHYCAPAGFAILKNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNG 207
Db 208 ACPKVSFEPIPIHYCAPAGFAILKCKDKKFKNGKGPCTNVSTVQCTHGIRPVVSTQLLNG 267
Qy 208 SLAEEGVWIRSENFTDNAKTIIVQLKESVEINCTRPNNNRKSTITIGPGRAFYATGDIIG 267
Db 268 SLAEEVWIRSANFADNAKVIIVQLNESVEINCTRPNNNRKSIHIGPGRAFYTGTGEIIG 327
Qy 268 DIRQAHCNISGEKWNNTLKQIVTKLOAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYC 327
Db 328 DIRQAHCNLSRAKWNNTLNKIVIKLREQFGNKTIVFKHSSGGDPEIVTHSFNCGGEFFYC 387
Qy 328 NSTQLFNSTWNNTIGPNNNT--NGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNI 385
Db 388 NSTQLFNSTWNVTESNNTVENNTITLPCRICKQIINRWQEVGRAMYAPPPIRGQIRCSSNI 447
Qy 386 TGLLLTRDGGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQ 445
Db 448 TGLLLTRDGGPE-DNKTEVFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQ 506
Qy 446 EKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL 505
Db 507 EKRAVGIGAVFLGFLGAAGSTMGAAAMTLTVQARLLSGIVQQNNLLRAIEAQOHL 566
Qy 506 TVMGK-LQARVLAVERYLKDQQLGIWCGSGKLICTTAVPWNASWSNKSOLDQIWNMTW 564
Db 567 TVMGKQLQARVLAVERYLRDQQLGIWCGSGKLICTTAVPWNASWSNKSOLDQIWNMTW 626
Qy 565 MEWERIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 627 IEWDREINNYTSIIYSLIEESQOQEKNEQELLELDKWASLWNWFDITKWLWYIK 681

RESULT 6

US-08-388-809-6
; Sequence 6, Application US/08388809
; Patent No. 5576000
; GENERAL INFORMATION:
; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
; APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
; APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
; APPLICANT: GARTNER, SUZANNE
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:

Qy 121 DR-----GEIKNCSFKVGAG-----KLINCNTSVITQ 147
Db 148 SRGMVGGGEMKCNCFNITTNIRGKVQKEYALFYKLDIAPIDNNSNNRYRLISCNTSVITQ 207
Qy 148 ACPKVSPEPIPIHYCAPAGFAILKCNDDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLLNG 207
Db 208 ACPKVSPEPIPIHYCAPAGFAILKCKKFKNGKGPCTNVSTVQCTHGIRPVVSTQLLLNG 267
Qy 208 SLAEEGVVIRSENFTDNAKTIIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIG 267
Db 268 SLAEEVVIRSANFADNAKVIIIVQLNESVEINCTRPNNNTRKSIHIGPGRAYTGTGEIIG 327
Qy 268 DIRQAHCNISGEKWNNTLKQIVTKLOAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYC 327
Db 328 DIRQAHCNLSRAKWNNTLNKIVIKLREQFGNKTIVFKHSSGGDPEIVTHSFNCGGEFFYC 387
Qy 328 NSTQFNSTWNNTIGPNNT--NGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSNI 385
Db 388 NSTQFNSTWNVTESNNTVENNTITLPCRIKQIINMWQEVGRAMYAPPPIRGQIRCSNI 447
Qy 386 TGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQR 445
Db 448 TGLLLTRDGGPE-DNKTEVFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQR 506
Qy 446 EKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQL 505
Db 507 EKRAVGIGAVFLGFLGAAGSTMGAAMTLTVQARLLLSGIVQQNNLLRAIEAQHLLQL 566
Qy 506 TVWGIK-LQARVLAVERYLKDQQLLGIGWCSGKLICTTAVPWNASWSNKSLDQIWNMTW 564
Db 567 TVWGIKQLQARVLAVERYLRDQQLLGIGWCSGKLICTTAVPWNASWSNKSLSNKIWDNMTW 626
Qy 565 MEWEREIDNYTNLIYTLIEESQNOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 627 IEWDREINNYTSIIYSLIEESQNOQEKNEQELLELDKWASLWNWFDITKWLWYIK 681

RESULT 8
US-09-124-900-9
; Sequence 9, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124,900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (79)..(184)
; NAME/KEY: PEPTIDE
; LOCATION: (326)..(400)
US-09-124-900-9

Query Match 85.7%; Score 2851; DB 3; Length 856;

Best Local Similarity 83.4%; Pred. No. 2.8e-226;
Matches 548; Conservative 32; Mismatches 37; Indels 40; Gaps 10;
Qy 1 SAVEKLWTVVYGVPMKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIULEN 60
Db 29 SATEKLWTVVYGVPMKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLVN 88
Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSGRMI 119
Db 89 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSGRMI 148
Qy 120 MDRGEIKNCSFKVGA---GK-----LINCNTSVITQACPKV 152
Db 149 MEKGEIKNCSFNISTSRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVITQACPKV 208
Qy 153 SPEPIPIHYCAPAGFAILKCNDDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 212
Db 209 SPEPIPIHYCAPAGFAILKCNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 268
Qy 213 GVVRSENFTDNAKTIIIVQLKESVEINCTRPNNNTRKSIITI--GPGRAFYATGDIIGDIR 270
Db 269 EVVIRSANFTDNAKTIIIVQLNQSVEINCTRPNNNTRKSIIRIQRGPGRAYFTIGK-IGNMR 327
Qy 271 QAHCNISGEKWNNTLKQIVTKLOAQFG-NKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNS 329
Db 328 QAHCNISRAKWNNTLKQIDSKLREQFGNKTIIIFKQSSGGDPEIVTHSFNCGGEFFYCNS 387
Qy 330 TQLFNSTWNN---TIGPNNTNG--TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS 383
Db 388 TQLFNSTWFNSTWSTKGSNNTGSDTITLPCRIKQIINMWQEVGKAMYAPPISGQIRCSS 447
Qy 384 NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVV 443
Db 448 NITGLLLTRDGGNS--NNESEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVV 506
Qy 444 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 503
Db 507 QREKRAVGIGALFLGFLGAAGSTMGAAMTLTVQARQLLSGIVQQNNLLRAIEAQHLL 566
Qy 504 QLTWGIK-LQARVLAVERYLKDQQLLGIGWCSGKLICTTAVPWNASWSNKSLDQIWNMM 562
Db 567 QLTWGIKQLQARILAVERYLKDQQLLGIGWCSGKLICTTAVPWNASWSNKSLEQIWNMM 626
Qy 563 TWMEWEREIDNYTNLIYTLIEESQNOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 627 TWMEWDREINNYTSLIHSLEESQNOQEKNEQELLELDKWASLWNWFNITNWLWYIK 683

RESULT 9
US-08-463-210-11
; Sequence 11, Application US/08463210
; Patent No. 6001977
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAI, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/463,210
; APPLICATION NUMBER: US/08/463,210
; FILING DATE: 05-JUN-1995

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,240A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/956,483
FILING DATE: 31-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 861 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Protein
LOCATION: 1..831
US-08-472-240A-7

Query Match 85.6%; Score 2848.5; DB 3; Length 861;
Best Local Similarity 82.8%; Pred. No. 4.5e-226;
Matches 548; Conservative 30; Mismatches 39; Indels 45; Gaps 10;

Qy	1	SAVEKLVTVYVGVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN	60
Db	29	SATEKLVTVYVGVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVN	88
Qy	61	VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWE	119
Db	89	VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVSLKCTDLGNATNTNSSNTSS	148
Qy	120	-----MDRGEIKNCSFKVGA---GK-----LINCNTSVITQ	147
Db	149	SGEMMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVITQ	208
Qy	148	ACPKVSFEPIPIHYCAPAGFAILKCNDDKFNKSGPCTNVSTVQCTHGIRPVVSTQLLLNG	207
Db	209	ACPKVSFEPIPIHYCAPAGFAILKCNKNTFNGTGCTNVSTVQCTHGIRPVVSTQLLLNG	268
Qy	208	SLAEEGVVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIIT--GPGRAFATGDI	265
Db	269	SLAEEVVIRSANFTDNAKTIIVQLNQSVINCTRPNNNTRKSIIRGPGRAFVTIGK-	327
Qy	266	IGDIRQAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIVFKQSSGGDPEIVMHSFNCGGEF	324
Db	328	IGNMRQAHCNISRAKNWATLKQIASKLRFQGNKNTIIFKQSSGGDPEIVTHSFNCGGEF	387
Qy	325	FYCNSTQLFNSTWNN---TIGPNNTNG--TITLPCRKQIINRWQEVGKAMYAPPPIRGQ	378
Db	388	FYCNSTQLFNSTWFSNTWSTEGSNNTSGSDTITLPCRKQFINMWQEVGKAMYAPPISGQ	447
Qy	379	IRCSSNITGLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTKA	438

Db	448	IRCSSNITGLLTRDGGNN--NNGSEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTKA	506
Qy	439	KRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQONNLLRAIEA	498
Db	507	KRRVVQREKRAVGIGALFLGFLGAAGSTMGARSMTLTVQARQLLSGIVQQONNLLRAIEA	566
Qy	499	QQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGIWGCSCGLICTTAVPWNASWSNKSLDQ	557
Db	567	QQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSCGLICTTAVPWNASWSNKSLEQ	626
Qy	558	IWNNTWMEWEREIDNYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWY	617
Db	627	IWNNTWMEWDREINNYTSLIHSIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWY	686
Qy	618	IK 619	
Db	687	IK 688	

Search completed: February 25, 2004, 14:20:12
Job time : 20.1051 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: February 25, 2004, 14:18:41 ; Search time 32.2957 Seconds
(without alignments)
4047.104 Million cell updates/sec

Title: US-09-891-609A-4
Perfect score: 3327
Sequence: 1 SAVEKLWVTYYGVPVWKEA.....DKWASLWNWFDISKWLWYIK 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3327	100.0	619	9 US-09-891-609-4	Sequence 4, Appli
2	3327	100.0	646	9 US-09-891-609-2	Sequence 2, Appli
3	3275	98.4	842	14 US-10-190-435-2	Sequence 2, Appli
4	3275	98.4	842	14 US-10-241-009-2	Sequence 2, Appli
5	3275	98.4	842	14 US-10-190-434B-2	Sequence 2, Appli
6	3275	98.4	842	14 US-10-190-305A-2	Sequence 2, Appli
7	3275	98.4	847	9 US-09-476-242-2	Sequence 2, Appli
8	2862	86.0	643	14 US-10-032-162-13	Sequence 13, Appl
9	2848.5	85.6	861	14 US-10-026-741-103	Sequence 103, App
10	2834	85.2	856	9 US-09-476-242-1	Sequence 1, Appli
11	2819	84.7	856	14 US-10-196-515-11	Sequence 11, Appl
12	2813.5	84.6	868	9 US-09-938-406-1	Sequence 1, Appli
13	2797.5	84.1	726	14 US-10-196-515-3	Sequence 3, Appli
14	2793.5	84.0	759	14 US-10-196-515-12	Sequence 12, Appl
15	2771	83.3	579	14 US-10-032-162-15	Sequence 15, Appl

16	2758	82.9	625	14	US-10-032-162-17	Sequence 17, Appl
17	2642.5	79.4	860	14	US-10-190-435-6	Sequence 6, Appli
18	2642.5	79.4	860	14	US-10-241-009-6	Sequence 6, Appli
19	2642.5	79.4	860	14	US-10-190-434B-6	Sequence 6, Appli
20	2642.5	79.4	860	14	US-10-190-305A-6	Sequence 6, Appli
21	2637.5	79.3	853	13	US-10-003-035-33	Sequence 33, Appl
22	2637.5	79.3	853	14	US-10-286-332A-33	Sequence 33, Appl
23	2637.5	79.3	853	15	US-10-280-915-33	Sequence 33, Appl
24	2588	77.8	870	14	US-10-190-435-147	Sequence 147, App
25	2585	77.7	867	14	US-10-190-435-3	Sequence 3, Appli
26	2585	77.7	867	14	US-10-190-435-126	Sequence 126, App
27	2585	77.7	867	14	US-10-241-009-3	Sequence 3, Appli
28	2585	77.7	867	14	US-10-190-434B-3	Sequence 3, Appli
29	2585	77.7	867	14	US-10-190-305A-3	Sequence 3, Appli
30	2582.5	77.6	865	14	US-10-190-435-140	Sequence 140, App
31	2579	77.5	858	14	US-10-190-435-150	Sequence 150, App
32	2573.5	77.4	855	14	US-10-190-435-144	Sequence 144, App
33	2570	77.2	862	14	US-10-190-435-141	Sequence 141, App
34	2566.5	77.1	870	14	US-10-190-435-127	Sequence 127, App
35	2564	77.1	869	14	US-10-190-435-4	Sequence 4, Appli
36	2564	77.1	869	14	US-10-241-009-4	Sequence 4, Appli
37	2564	77.1	869	14	US-10-190-434B-4	Sequence 4, Appli
38	2564	77.1	869	14	US-10-190-305A-4	Sequence 4, Appli
39	2558	76.9	861	14	US-10-190-435-139	Sequence 139, App
40	2555.5	76.8	845	14	US-10-190-435-129	Sequence 129, App
41	2555.5	76.8	845	14	US-10-190-435-130	Sequence 130, App
42	2554.5	76.8	803	14	US-10-190-435-134	Sequence 134, App
43	2551.5	76.7	803	14	US-10-190-435-135	Sequence 135, App
44	2547.5	76.6	857	14	US-10-190-435-138	Sequence 138, App
45	2539.5	76.3	845	14	US-10-190-435-143	Sequence 143, App

ALIGNMENTS

RESULT 1
US-09-891-609-4
; Sequence 4, Application US/09891609
; Patent No. US20020127238A1
; GENERAL INFORMATION:
; APPLICANT: Stamatos, Leonidas
; TITLE OF INVENTION: HIV-1 VACCINES AND SCREENING METHODS THEREFOR
; FILE REFERENCE: 2570-1-002N
; CURRENT APPLICATION NUMBER: US/09/891,609
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,608
; PRIOR FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-891-609-4

Query Match		100.0%;	Score 3327;	DB 9;	Length 619;
Best Local Similarity		100.0%;	Pred. No. 2.8e-296;		
Matches 619;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	SAVEKLWVTYYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEI	VLEN 60		
Db	1	SAVEKLWVTYYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEI	VLEN 60		
Qy	61	VTEFNFMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNA	TNTKSSNWKEM 120		
Db	61	VTEFNFMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNA	TNTKSSNWKEM 120		
Qy	121	DRGEIKNCSFKVGAGKLINCNSTSVITQACPKVSFEPIPIHYCAPAG	FAILKCNCKKFN		
Db	121	DRGEIKNCSFKVGAGKLINCNSTSVITQACPKVSFEPIPIHYCAPAG	FAILKCNCKKFN		
Qy	181	GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEQGVIRSENFDTNAKTI	IIVOLKESVEIN		
Db	181	GPCTNVSTVQCTHGIRPVVSTQLLNGSLAEQGVIRSENFDTNAKTI	IIVOLKESVEIN		

Db 181 GPCNTVSTVQCTHGIRPVVSTQLLNGSLAEAGVVRSENFTDNAKTIIVQLKESVEINC 240

Qy 241 TRPNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNKT 300

Db 241 TRPNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNKT 300

Qy 301 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNTNGTITLPCRICKQII 360

Db 301 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNTNGTITLPCRICKQII 360

Qy 361 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 420

Db 361 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 420

Qy 421 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 480

Db 421 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 480

Qy 481 LLSGIVQQNNLLRAIEAQOHLQLTVWGIKLARVLAVERYLKDQQLGIWGCSGKLIC 540

Db 481 LLSGIVQQNNLLRAIEAQOHLQLTVWGIKLARVLAVERYLKDQQLGIWGCSGKLIC 540

Qy 541 TTAVPWNASWSNKSLDQIWNNTWMWEEREIDNYTNLIYTLIEESQNQQEKNEQELLELD 600

Db 541 TTAVPWNASWSNKSLDQIWNNTWMWEEREIDNYTNLIYTLIEESQNQQEKNEQELLELD 600

Qy 601 KWASLWNWFDISKWLWYIK 619

Db 601 KWASLWNWFDISKWLWYIK 619

RESULT 2

US-09-891-609-2

; Sequence 2, Application US/09891609

; Patent No. US20020127238A1

; GENERAL INFORMATION:

; APPLICANT: Stamatatos, Leondias

; TITLE OF INVENTION: HIV-1 VACCINES AND SCREENING METHODS THEREFOR

; FILE REFERENCE: 2570-1-002N

; CURRENT APPLICATION NUMBER: US/09/891,609

; CURRENT FILING DATE: 2001-06-26

; PRIOR APPLICATION NUMBER: 60/214,608

; PRIOR FILING DATE: 2000-06-27

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 646

; TYPE: PRT

; ORGANISM: Human immunodeficiency virus type 1

US-09-891-609-2

Query Match 100.0%; Score 3327; DB 9; Length 646;

Best Local Similarity 100.0%; Pred. No. 3e-296;

Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SAVEKLWVTVYVGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60

Db 28 SAVEKLWVTVYVGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 87

Qy 61 VTENFNWKNMVQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

Db 88 VTENFNWKNMVQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 147

Qy 121 DRGEIKNCSFKVGAGKLINCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNDDKKFNGS 180

Db 148 DRGEIKNCSFKVGAGKLINCNTSVITQACPVSFEPIPIHYCAPAGFAILKCNDDKKFNGS 207

Qy 181 GPCNTVSTVQCTHGIRPVVSTQLLNGSLAEAGVVRSENFTDNAKTIIVQLKESVEINC 240

Db 208 GPCNTVSTVQCTHGIRPVVSTQLLNGSLAEAGVVRSENFTDNAKTIIVQLKESVEINC 267

Qy 241 TRPNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNKT 300

Db 268 TRPNNTRKSIITIGPGRAFYATGDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNKT 327

Qy 301 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNTNGTITLPCRICKQII 360

Db 328 IVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNSTWNNTIGPNNTNGTITLPCRICKQII 387

Qy 361 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 420

Db 388 NRQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSEL 447

Qy 421 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 480

Db 448 YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 507

Qy 481 LLSGIVQQNNLLRAIEAQOHLQLTVWGIKLARVLAVERYLKDQQLGIWGCSGKLIC 540

Db 508 LLSGIVQQNNLLRAIEAQOHLQLTVWGIKLARVLAVERYLKDQQLGIWGCSGKLIC 567

Qy 541 TTAVPWNASWSNKSLDQIWNNTWMWEEREIDNYTNLIYTLIEESQNQQEKNEQELLELD 600

Db 568 TTAVPWNASWSNKSLDQIWNNTWMWEEREIDNYTNLIYTLIEESQNQQEKNEQELLELD 627

Qy 601 KWASLWNWFDISKWLWYIK 619

Db 628 KWASLWNWFDISKWLWYIK 646

RESULT 3

US-10-190-435-2

; Sequence 2, Application US/10190435

; Publication No. US20030143248A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MEGEDE, Jan

; APPLICANT: BARNETT, Susan W.

; APPLICANT: LIAN, Ying

; APPLICANT: ENGELBRECHT, Susan

; APPLICANT: VAN RENSBURG, Estrelita J.

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C

; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

; FILE REFERENCE: PP18133.003 / 2302-18133

; CURRENT APPLICATION NUMBER: US/10/190,435

; CURRENT FILING DATE: 2002-12-30

; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 842

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SF162

US-10-190-435-2

Query Match 98.4%; Score 3275; DB 14; Length 842;

Best Local Similarity 95.2%; Pred. No. 2.6e-291;

Matches 616; Conservative 0; Mismatches 3; Indels 28; Gaps 2;

Qy 1 SAVEKLWVTVYVGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60

Db 23 SAVEKLWVTVYVGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 82

Qy 61 VTENFNWKNMVQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

Db 83 VTENFNWKNMVQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 142

Qy 121 DRGEIKNCSFKVGAG-----KLNICNTSVITQACPVS 153

Db 143 DRGEIKNCSFKVTTTSIRNMQKEYALFYKLDVVPIDNDNTSYKLINCNTSVITQACPVS 202

Qy 154 FEPIPIHYCAPAGFAILKCNDDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAG 213

Db 203 FEPIPIHYCAPAGFAILKCNDDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEAG 262

Qy 214 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273

Db 263 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 322
Qy 274 CNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 333
Db 323 CNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 382
Qy 334 NSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 393
Db 383 NSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 442
Qy 394 GGKEISNTTEIFRPGGDMRDNRSELYKYKVKIEPLGVAPTAKARRVVQREKRAVTLG 453
Db 443 GGKEISNTTEIFRPGGDMRDNRSELYKYKVKIEPLGVAPTAKARRVVQREKRAVTLG 502
Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
Db 503 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQL 562
Qy 513 QARVLAVERYLKDQQLLGIWCGSKLICTTAVPWNASWSNKSLLDQIWNNTMTMEWEREID 572
Db 563 QARVLAVERYLKDQQLLGIWCGSKLICTTAVPWNASWSNKSLLDQIWNNTMTMEWEREID 622
Qy 573 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 623 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 4
US-10-241-009-2
; Sequence 2, Application US/10241009
; Publication No. US20030170614A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.21
; CURRENT APPLICATION NUMBER: US/10/241,009
; CURRENT FILING DATE: 2002-12-13
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SF162
US-10-241-009-2

Query Match 98.4%; Score 3275; DB 14; Length 842;
Best Local Similarity 95.2%; Pred. No. 2.6e-291;
Matches 616; Conservative 0; Mismatches 3; Indels 28; Gaps 2;
Qy 1 SAVEKLWVTVYVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60
Db 23 SAVEKLWVTVYVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 82
Qy 61 VTENFNWKNMNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
Db 83 VTENFNWKNMNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 142
Qy 121 DRGEIKNCSFKVAG-----KLINCNTSVITQACPKVS 153
Db 143 DRGEIKNCSFKVTTTSIRNKMCKEYALFYKLDVVPIDNDNTSYKLINCNTSVITQACPKVS 202
Qy 154 FEPIPIHYCAPAGFAILKCNCKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 213
Db 203 FEPIPIHYCAPAGFAILKCNCKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 262
Qy 214 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273

Db 263 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 322
Qy 274 CNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 333
Db 323 CNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 382
Qy 334 NSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 393
Db 383 NSTWNNTIGPNNNTGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 442
Qy 394 GGKEISNTTEIFRPGGDMRDNRSELYKYKVKIEPLGVAPTAKARRVVQREKRAVTLG 453
Db 443 GGKEISNTTEIFRPGGDMRDNRSELYKYKVKIEPLGVAPTAKARRVVQREKRAVTLG 502
Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
Db 503 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQL 562
Qy 513 QARVLAVERYLKDQQLLGIWCGSKLICTTAVPWNASWSNKSLLDQIWNNTMTMEWEREID 572
Db 563 QARVLAVERYLKDQQLLGIWCGSKLICTTAVPWNASWSNKSLLDQIWNNTMTMEWEREID 622
Qy 573 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 623 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK 669

RESULT 5
US-10-190-434B-2
; Sequence 2, Application US/10190434B
; Publication No. US20030194800A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2300-1621.20
; CURRENT APPLICATION NUMBER: US/10/190,434B
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SF162
US-10-190-434B-2

Query Match 98.4%; Score 3275; DB 14; Length 842;
Best Local Similarity 95.2%; Pred. No. 2.6e-291;
Matches 616; Conservative 0; Mismatches 3; Indels 28; Gaps 2;
Qy 1 SAVEKLWVTVYVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60
Db 23 SAVEKLWVTVYVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 82
Qy 61 VTENFNWKNMNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
Db 83 VTENFNWKNMNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 142
Qy 121 DRGEIKNCSFKVAG-----KLINCNTSVITQACPKVS 153
Db 143 DRGEIKNCSFKVTTTSIRNKMCKEYALFYKLDVVPIDNDNTSYKLINCNTSVITQACPKVS 202
Qy 154 FEPIPIHYCAPAGFAILKCNCKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 213
Db 203 FEPIPIHYCAPAGFAILKCNCKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 262
Qy 214 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273
Db 263 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 322

Qy	274	CNISGEKWNNTLKQIVTKLQAFGNKTI	VFVKQSSGGDPEIVMHSFNCGEFFYC	NSTQLF	333
Db	323	CNISGEKWNNTLKQIVTKLQAFGNKTI	VFVKQSSGGDPEIVMHSFNCGEFFYC	NSTQLF	382
Qy	334	NSTWNNTIGPNNTNGTITLPCR	IKQIINRWQEVGKAMYAPPIRG	QIRCSSNITGLLLTRD	393
Db	383	NSTWNNTIGPNNTNGTITLPCR	IKQIINRWQEVGKAMYAPPIRG	QIRCSSNITGLLLTRD	442
Qy	394	GGKEISNTTEIFRPGGGDMRDNR	SELYKYKVKVIEPLGVAPT	KAKRRVVQREKRAVTLG	453
Db	443	GGKEISNTTEIFRPGGGDMRDNR	SELYKYKVKVIEPLGVAPT	KAKRRVVQREKRAVTLG	502
Qy	454	AMFLGFLGAAGSTMGARSLTLTVQ	ARQLLSGIVQQONLLRAIEAQ	HLLQLTVMGIK-L	512
Db	503	AMFLGFLGAAGSTMGARSLTLTVQ	ARQLLSGIVQQONLLRAIEAQ	HLLQLTVMGIKQL	562
Qy	513	QARVLAVERYLKDQQLLGWCGSK	LICTTAVPWNASWSNKSLDQI	WNNTWMEWEREID	572
Db	563	QARVLAVERYLKDQQLLGWCGSK	LICTTAVPWNASWSNKSLDQI	WNNTWMEWEREID	622
Qy	573	NYTNLIYTLIEESQNOEKNQE	ELLELDKWSLWNWFDISK	WLWYIK	619
Db	623	NYTNLIYTLIEESQNOEKNQE	ELLELDKWSLWNWFDISK	WLWYIK	669

Qy	274	CNISGEKNWNTLKQIVTKLQAFGNKTI	VF	KOSSGGDPEI	VMHSFNCGGEFFYC	NSTQLF	333
Db	323	CNISGEKNWNTLKQIVTKLQAFGNKTI	VF	KOSSGGDPEI	VMHSFNCGGEFFYC	NSTQLF	382
Qy	334	NSTWNNTIGPNNTNGTITLPCRIKQI	IN	RWQEVGKAMY	APP	IRGQIRCSSNITGLLLTRD	393
Db	383	NSTWNNTIGPNNTNGTITLPCRIKQI	IN	RWQEVGKAMY	APP	IRGQIRCSSNITGLLLTRD	442
Qy	394	GGKEISNTTIEFRPGGGMRDNWRSELY	KY	KVKIEPLGVAPT	KAKRRVVQREKRAVTLG	453	
Db	443	GGKEISNTTIEFRPGGGMRDNWRSELY	KY	KVKIEPLGVAPT	KAKRRVVQREKRAVTLG	502	
Qy	454	AMFLGFLGAAGSTMGARSLTLTVQARQL	LSG	IVQQNNLLRAIEA	QQHLLQLTVWG	IK-L	512
Db	503	AMFLGFLGAAGSTMGARSLTLTVQARQL	LSG	IVQQNNLLRAIEA	QQHLLQLTVWG	IKQL	562
Qy	513	QARVLAVERYLKDDQLLGIWGC	SGKLI	CTTAVPWNASWSNKS	LDQIWNMTWMEW	EREID	572
Db	563	QARVLAVERYLKDDQLLGIWGC	SGKLI	CTTAVPWNASWSNKS	LDQIWNMTWMEW	EREID	622
Qy	573	NYTNLIYTLIEESQOQEKNEQELLE	LDKWASLWNWFDISKW	LWYIK	619		
Db	623	NYTNLIYTLIEESQOQEKNEQELLE	LDKWASLWNWFDISKW	LWYIK	669		

Qy 334 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 393
Db 388 NSTWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 447
Qy 394 GGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLG 453
Db 448 GGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLG 507
Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
Db 508 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQL 567
Qy 513 QARVLAVERYLKDQQLLGWCGSKLICTTAVPWNASWSNKSLDQIWNNTMTMEWEREID 572
Db 568 QARVLAVERYLKDQQLLGWCGSKLICTTAVPWNASWSNKSLDQIWNNTMTMEWEREID 627
Qy 573 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNNWFDISKWLWYIK 619
Db 628 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNNWFDISKWLWYIK 674

RESULT 8

US-10-032-162-13
; Sequence 13, Application US/10032162
; Publication No. US20030052839A1
; GENERAL INFORMATION:
; APPLICANT: BINLEY, JAMES M
; APPLICANT: SCHUELKE, NORBERT
; APPLICANT: OLSON, WILLIAM C
; APPLICANT: PAUL, MADDON J
; APPLICANT: JOHN, MOORE P
; TITLE OF INVENTION: STABILIZED VIRAL ENVELOPE PROTEINS AND USES THEREOF
; FILE REFERENCE: 2048/59331az
; CURRENT APPLICATION NUMBER: US/10/032,162
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 09/602,864
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-032-162-13

Query Match 86.0%; Score 2862; DB 14; Length 643;
Best Local Similarity 84.5%; Pred.No. 1.4e-253;
Matches 545; Conservative 25; Mismatches 43; Indels 32; Gaps 6;

Qy 3 VEKLWTVYVYGVVPWKEATTLTFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLENT 62
Db 1 VEKLWTVYVYGVVPWKEATTLTFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLENT 60
Qy 63 ENFNWKNMNVEMQHEDIISLWDSILKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEMDR 122
Db 61 EHFNWKNMNVEMQMEDIISLWDSILKPCVKLTPLCVTLNCKDV-NATNTTNDSEGTMER 119
Qy 123 GEIKNCSPKVGAG-----KLINCNTSVITQACPKVSFE 155
Db 120 GEIKNCSEFNITTSIRDEVQKEYALFYKLDVVPIDNNNTSYRLISCDTSVITQACPKISFE 179
Qy 156 PIPHYCAPAGFAILKCNKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEGVV 215
Db 180 PIPHYCAPAGFAILKCNCKDTFNGKGCKNVSTVQCTHGIRPVVSTQLLLNGSLAEVEV 239
Qy 216 IRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFATGDIIGDIRQAHCN 275
Db 240 IRSDFNTNAKTIIVQLKESVEINCTRPNNNTRKSIHIGPGRAFYTGTGEIIGDIRQAHCN 299
Qy 276 ISGEKNWNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNS 335
Db 300 ISRAKWNDTLKQIVIKLREQFENKTIVFNHSSGGDPEIVMHSFNCEGEFFYCNSTQLFNS 359

Qy 336 TW-NNTIGPNNNTG-TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 393
Db 360 TWNNTGEGSNNTGNTITLPCRIKQIINMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD 419
Qy 394 GGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLG 453
Db 420 GGIN-ENGTEIFRPGGDMRDNRSEFYKYKVVKIEPLGVAPTCKRRVVQREKRAVGIG 478
Qy 454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
Db 479 AVFLGFLGAAGSTMGAASMTLTVQARLLSGIVQQNNLLRAIEAQQRMQLQTVWGIKQL 538
Qy 513 QARVLAVERYLKDQQLLGWCGSKLICTTAVPWNASWSNKSLDQIWNNTMTMEWEREID 572
Db 539 QARVLAVERYLGDQQLLGWCGSKLICTTAVPWNASWSNKSLDRIWNNMTMTMEWEREID 598
Qy 573 NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNNWFDISKWLWY 617
Db 599 NYTSEIYTLIEESQNQQEKNEQELLELDKWASLWNNWFDITNLWY 643

RESULT 9

US-10-026-741-103
; Sequence 103, Application US/10026741
; Publication No. US20030049604A1
; GENERAL INFORMATION:
; APPLICANT: CHARNEAU, PIERRE
; CLAVEL, FRANCOISE
; BORMAN, ANDREW
; QUILLET, CAROLINE
; GUETARD, DENISE
; MONTAGNIER, LUC
; DONJON DE SAINT-MARTIN, JACQUELINE
; COHEN, JAUQUES
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR
SUBTYPE) ANTIGENS
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &
; Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/026,741
; FILING DATE: 27-Dec-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/817,441
; FILING DATE: 31-AUG-1998
; APPLICATION NUMBER: PCT/FR 95/01391
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: FR 9412554
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: FR 9502526
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03260.6005-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 amino acids
; TYPE: amino acid
;

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-10-026-741-103

Query Match      85.6%; Score 2848.5; DB 14; Length 861;
Best Local Similarity 82.8%; Pred. No. 3.8e-252;
Matches 548; Conservative 30; Mismatches 39; Indels 45; Gaps 10;

Qy 1 SAVEKLVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 29 SATEKLVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLN 88

Qy 61 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTKSSNWKE- 119
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 89 VTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLGNATNTSSNTSS 148

Qy 120 -----MDRGEIKNCSEFKVGA---GK-----LINCNTSVITQ 147
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 149 SGEMMEKGEIKNCSEFNISTIRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVITQ 208

Qy 148 ACPKVSPEPIHYCAPAGFAILKNDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNG 207
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 209 ACPKVSPEPIHYCAPAGFAILKCNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLNG 268

Qy 208 SLAEEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSITI--GPGRAFYATGDI 265
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 269 SLAEEVVIRSENFDTNAKTIIVQLNQSVVEINCTRPNNNTRKSIRIQRGPGRAFTIGK- 327

Qy 266 IGDIRQAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIIVFKQSSGGDPEIVMHSFNCGGEF 324
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 IGMRQAHCNISRAKNATLKQIASKLREQFGNKNKTIIFKQSSGGDPEIVTHSFNCGGEF 387

Qy 325 FYCNSTQLFNSTWNN---TIGPNNNG--TITLPCRKQIINRWQEVGKAMYAPPPIRGQ 378
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 388 FYCNSTQLFNSTWFNSTWSTEGSNNTGSDTITLPCRKQFINNMWQEVGKAMYAPPISGQ 447

Qy 379 IRCSSNITGLLLTRDGGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKA 438
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 448 IRCSSNITGLLLTRDGGNN--NNGSEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTKA 506

Qy 439 KRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEA 498
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 507 KRRVVQREKRAVGIGALFLGFLGAAGSTMGARSMTLTVQARQLLSGIVQQNNLLRAIEA 566

Qy 499 QHLLQLTVWGIK-LQARVLAVERYLKDDQLLGIWCGSGKLICTTAVPWNASWSNKSLEQ 557
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 567 QHLLQLTVWGIKQLQARILAVERYLKDDQLLGIWCGSGKLICTTAVPWNASWSNKSLEQ 626

Qy 558 IWNMTWMEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWY 617
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 627 IWNMTWMEWDREINNYTSLIHSLEESQOQEKNEQELLELDKWASLWNWFNITNWLWY 686

Qy 618 IK 619
   ||
Db 687 IK 688

RESULT 10
US-09-476-242-1
; Sequence 1, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-476-242-1

Query Match      85.2%; Score 2834; DB 9; Length 856;
Best Local Similarity 82.6%; Pred. No. 8e-251;
Matches 543; Conservative 33; Mismatches 41; Indels 40; Gaps 9;

Qy 1 SAVEKLVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 29 SATEKLVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLN 88

Qy 61 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTKSSNWKE- 119
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 89 VTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMI 148

Qy 120 MDRGEIKNCSEFKVGA-----KLINCNTSVITQACPKV 152
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 149 MEKGEIKNCSEFNISTIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKV 208

Qy 153 SPEPIPIHYCAPAGFAILKNDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 212
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 209 SPEPIPIHYCAPAGFAILKCNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 268

Qy 213 GVVRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSITI--GPGRAFYATGDIIGDIR 270
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 269 EVVIRSVNFDTNAKTIIVQLNTSVEINCTRPNNNTRKIRIQRGPGRAFTIGK-IGNMR 327

Qy 271 QAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCN 329
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 QAHCNISRAKNNTLKQIASKLREQFGNKNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCN 387

Qy 330 TQLFNSTWNN---TIGPNNNG--TITLPCRKQIINRWQEVGKAMYAPPPIRGQIRCSS 383
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 388 TQLFNSTWFNSTWSTEGSNNTGSDTITLPCRKQIINMWQVKGKAMYAPPISGQIRCSS 447

Qy 384 NITGLLLTRDGGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV 443
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 448 NITGLLLTRDGGNS--NNESEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV 506

Qy 444 QREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLL 503
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 507 QREKRAVGIGALFLGFLGAAGSTMGAASMTLTQARQLLSGIVQQNNLLRAIEAQHLL 566

Qy 504 QLTWGIK-LQARVLAVERYLKDDQLLGIWCGSGKLICTTAVPWNASWSNKSLEQIWNHT 562
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 567 QLTWGIKQLQARILAVERYLKDDQLLGIWCGSGKLICTTAVPWNASWSNKSLEQIWNHT 626

Qy 563 TWMEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 627 TWMEWDREINNYTSLIHSLEESQOQEKNEQELLELDKWASLWNWFNITNWLWYIK 683

RESULT 11
US-10-196-515-11
; Sequence 11, Application US/10196515
; Publication No. US20030091594A1
; GENERAL INFORMATION:
; APPLICANT: HOXIE, James A.
; APPLICANT: LABRANCHE, Celia C.
; APPLICANT: DOMS, Robert W.
; APPLICANT: HOFFMAN, Trevor L.
; TITLE OF INVENTION: CD4-INDEPENDENT HIV ENVELOPE PROTEINS AS VACCINES AND
; TITLE OF INVENTION: THERAPEUTICS
; FILE REFERENCE: Hoxie 9596-104U1 (0282)
; CURRENT APPLICATION NUMBER: US/10/196,515
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/337,387
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: US 09/317,556
; PRIOR FILING DATE: 1999-05-24
```



```
; TITLE OF INVENTION: CD4-INDEPENDENT HIV ENVELOPE PROTEINS AS VACCINES AND
; TITLE OF INVENTION: THERAPEUTICS
; FILE REFERENCE: Hoxie 9596-104U1 (0282)
; CURRENT APPLICATION NUMBER: US/10/196,515
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/337,387
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: US 09/317,556
; PRIOR FILING DATE: 1999-05-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-196-515-3
```

```
Query Match      84.1%; Score 2797.5; DB 14; Length 726;
Best Local Similarity 81.8%; Pred. No. 1.4e-247;
Matches 534; Conservative 37; Mismatches 45; Indels 37; Gaps 9;
```

```
Qy      1 SAVEKLVVTYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60
Db      29 NATEKLVVTYYGVPVWKEATTTLFCASDAKAYETEVHNVWATHACVPTDPNPQEVVLVN 88

Qy      61 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKE- 119
Db      89 VTENFNMMKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSGSRMI 148

Qy      120 MDRGEIKNCSFKVGAGK-----LINCNTSVITQACPKV 152
Db      149 MEKGEIKNCSFNISTSKRSKVKKEYAFFYKLDIIPIDNDPTSYYTLTSCNTSVITQACPKV 208

Qy      153 SPEPIPIHYCAPAGFAILKCNCKFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEE 212
Db      209 SFEPPIPIHYCAPAGFAILKCNKNTFNGTGCTNVSTVQCTHGIRPVVSTQLLLNGSLAEE 268

Qy      213 GVVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITI--GPGRAFYATGDIIGDIR 270
Db      269 EVVIRSVNFTDNAKTIIVQLNTSVEINCTKPNNNTRKRIRIHRGPGRAFTVVGK-IGNMR 327

Qy      271 QAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIIVFKOSSGGDPEIVMHSFNCGGEFFYCNS 329
Db      328 QAHCNISRAKWSNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCKS 387

Qy      330 TQLFNSTWNNTIGPNNNG--TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITG 387
Db      388 TQLFNSTW-STKGSNNTEGSDTITLPCRIKQVINMWQEVGKAMYAPPISGQIRCSSNITG 446

Qy      388 LLLTRDGGKEISNTTEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREK 447
Db      447 LLLTRDGGNS--NNESEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREK 505

Qy      448 RAVTLGAMFLGFLGAAGSTMGARSILTLTVQARQLLSGIVQQQNNLLRAIEAQHLLQLTV 507
Db      506 RAVGIGALFLGFLGAAGSTMGAASMALTVQARQSLSGIVQQQNNLLRAIEAQHLLQLTV 565

Qy      508 WGIK-LOARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLDQIWNNTWTME 566
Db      566 WGIKQLQARILAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLEQIWNNTWTME 625

Qy      567 WEREIDNYTNLIYTLIEESQNOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db      626 WDREINNYTSLIHSLIEESQIQEWMNEQELLELDKWASLWNWFNITNWLWYIK 678
```

```
RESULT 14
US-10-196-515-12
; Sequence 12, Application US/10196515
; Publication No. US20030091594A1
; GENERAL INFORMATION:
; APPLICANT: HOXIE, James A.
; APPLICANT: LABRANCHE, Celia C.
```

```
; APPLICANT: DOMS, Robert W.
; APPLICANT: HOFFMAN, Trevor L.
; TITLE OF INVENTION: CD4-INDEPENDENT HIV ENVELOPE PROTEINS AS VACCINES AND
; TITLE OF INVENTION: THERAPEUTICS
; FILE REFERENCE: Hoxie 9596-104U1 (0282)
; CURRENT APPLICATION NUMBER: US/10/196,515
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/337,387
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: US 09/317,556
; PRIOR FILING DATE: 1999-05-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 759
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-196-515-12
```

```
Query Match      84.0%; Score 2793.5; DB 14; Length 759;
Best Local Similarity 81.9%; Pred. No. 3.5e-247;
Matches 535; Conservative 36; Mismatches 45; Indels 37; Gaps 9;
```

```
Qy      1 SAVEKLVVTYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60
Db      29 NATEKLVVTYYGVPVWKEATTTLFCASDAKAYETEVHNVWATHACVPTDPNPQEVVLVN 88

Qy      61 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKE- 119
Db      89 VTENFNMMKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSGSRMI 148

Qy      120 MDRGEIKNCSFKVGAGK-----LINCNTSVITQACPKV 152
Db      149 MEKGEIKNCSFNISTSKRGKVKKEYAFFYKLDIIPIDNDPTSYYTLTSCNTSVITQACPKV 208

Qy      153 SFEPPIPIHYCAPAGFAILKCNCKFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEE 212
Db      209 SFEPPIPIHYCAPAGFAILKCNKNTFNGTGCTNVSTVQCTHGIRPVVSTQLLLNGSLAEE 268

Qy      213 GVVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITI--GPGRAFYATGDIIGDIR 270
Db      269 EVVIRSVNFTDNAKTIIVQLNTSVEINCTKPNNNTRKRIRIQRGPGRAFTVVGK-IGNMR 327

Qy      271 QAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIIVFKOSSGGDPEIVMHSFNCGGEFFYCNS 329
Db      328 QAHCNISRAKWSNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCKS 387

Qy      330 TQLFNSTWNNTIGPNNNG--TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITG 387
Db      388 TQLFNSTW-STKGSNNTEGSDTITLPCRIKQIINMWQKVEKAMYAPPISGQIRCSSNITG 446

Qy      388 LLLTRDGGKEISNTTEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREK 447
Db      447 LLLTRDGGNN--NNESEIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREK 505

Qy      448 RAVTLGAMFLGFLGAAGSTMGARSILTLTVQARQLLSGIVQQQNNLLRAIEAQHLLQLTV 507
Db      506 RAVGIGALFLGFLGAAGSTMGAASMALTVQARQSLSGIVQQQNNLLRAIEAQHLLQLTV 565

Qy      508 WGIK-LOARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLDQIWNNTWTME 566
Db      566 WGIKQLQARILAVERYLKDQQLLGIWCGSGKLICTTAVPWSASWSNKSLEQIWNNTWTME 625

Qy      567 WEREIDNYTNLIYTLIEESQNOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db      626 WDREINNYTSLIHSLIEESQNOQEMNEQELLELDKWASLWNWFIISWLWYIK 678
```

```
RESULT 15
US-10-032-162-15
; Sequence 15, Application US/10032162
; Publication No. US20030052839A1
; GENERAL INFORMATION:
```


; APPLICANT: BINLEY, JAMES M
; APPLICANT: SCHUELKE, NORBERT
; APPLICANT: OLSON, WILLIAM C
; APPLICANT: PAUL, MADDON J
; APPLICANT: JOHN, MOORE P
; TITLE OF INVENTION: STABILIZED VIRAL ENVELOPE PROTEINS AND USES THEREOF
; FILE REFERENCE: 2048/59331az
; CURRENT APPLICATION NUMBER: US/10/032,162
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: 09/602,864
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-032-162-15

Query Match		83.3%;	Score 2771;	DB 14;	Length 579;
Best Local Similarity		85.6%;	Pred. No. 2.7e-245;		
Matches 529;		Conservative 16;	Mismatches 31;	Indels 42;	Gaps 6;
Qy	3	VEKLWVTYYGVPVWKEATTTLCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLENT 62			
Db	1	VEKLWVTYYGVPVWKEATTTLCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLENT 60			
Qy	63	ENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEMDR 122			
Db	61	EHFNMWKNMVEQMEDIISLWDQSLKPCVKLTPLC----- 96			
Qy	123	GEIKNCSEFKVGAGKLINCNTSVITQACPKVSFEPIPIHYCAPAGFAILKNDKKFNGSGP 182			
Db	97	-----GAG-----CDTSVITQACPKISFEPIPIHYCAPAGFAILKNDKTFNGKGP 142			
Qy	183	CTNVSTVQCTHGIRPVVSTQLLNGSLAEAGVWIRSENFDTNAKTIIVQLKESVEINCTR 242			
Db	143	CKNVSTVQCTHGIRPVVSTQLLNGSLAEAEWIRSDNFNTNAKTIIVQLKESVEINCTR 202			
Qy	243	PNNTRKSIITGPGRAFYATGDIIGDIRQAHNCISGEKWNNTLKQIVTKLQAQFGNKTIV 302			
Db	203	PNNTRKSIHIGPGRAFYTTGEIIGDIRQAHNCISRAKWNNTLKQIVIKLREQFENKTIV 262			
Qy	303	EKQSSGGDEIVMHSFNCGGEFFYCNSTQLFNSWTW-NNTIGPNNNTNG-TITLPCRICKQII 360			
Db	263	FNHSSGGDEIVMHSFNCGGEFFYCNSTQLFNSWTWNNNTGSGNNTGNTITLPCRICKQII 322			
Qy	361	NRQVEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGGMDRDNWRSEL 420			
Db	323	NMQVEVGKAMYAPPPIRGQIRCSSNITGLLLTRDGGIN-ENGTEIFRPGGGMDRDNWRSEL 381			
Qy	421	YKYKVVKIEPLGVAPTAKRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQ 480			
Db	382	YKYKVVKIEPLGVAPTKRRVVQREKRAVGIGAVFLGFLGAAGSTMGAASMTLTVQARL 441			
Qy	481	LLSGIVQQNNLLRAIEAQOHLQLTWGKIK-LQARVLAVERYLKDQQLLGIWCGSGKLI 539			
Db	442	LLSGIVQQNNLLRAIEAQOQRLQLTWGKIKQLQARVLAVERYLGDQQLLGIWCGSGKLI 501			
Qy	540	CTTAVPWNASWSNKSQDIQNNMTWMEWEREIDNYTNLIYTLIEESONQOQKNEQELLEL 599			
Db	502	CCTAVPWNASWSNKSQDIQNNMTWMEWEREIDNYTSEIYTLIEESONQOQKNEQELLEL 561			
Qy	600	DKWASLWNWFDISKWLWY 617			
Db	562	DKWASLWNWFDITNWLWY 579			

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: February 25, 2004, 14:12:55 ; Search time 17.6158 Seconds
(without alignments)
3380.063 Million cell updates/sec

Title: US-09-891-609A-4
Perfect score: 3327
Sequence: 1 SAVEKLMWTVVYGVVWKEA.....DKWASLWNWFDISKWLWYIK 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2906	87.3	843	1 H44001	env polyprotein pr
2	2906	87.3	847	2 T09448	envelope glycoprot
3	2900.5	87.2	852	2 T12016	envelope glycoprot
4	2889	86.8	847	2 S13289	env protein - huma
5	2856	85.8	855	1 VCLJA2	env polyprotein pr
6	2851	85.7	856	1 VCLJH3	env polyprotein pr
7	2848.5	85.6	861	1 VCLJLV	env polyprotein pr
8	2839.5	85.3	851	2 S33985	env polyprotein -
9	2838	85.3	856	1 VCLJVL	env polyprotein pr
10	2836	85.2	861	1 VCLJSC	env polyprotein pr
11	2830	85.1	854	2 S13288	env protein - huma
12	2817	84.7	852	1 VCLJBR	env polyprotein -
13	2813.5	84.6	868	1 VCLJH4	env polyprotein -
14	2792.5	83.9	729	1 VCLJKB	env polyprotein pr
15	2792.5	83.9	861	1 VCLJKB	env polyprotein pr
16	2787.5	83.8	856	1 VCLJ3W	env polyprotein pr
17	2772.5	83.3	859	1 VCLJMN	env polyprotein pr
18	2593	77.9	855	1 VCLJZR	env polyprotein pr
19	2583.5	77.7	859	2 T01672	envelope polyprote
20	2571	77.3	853	2 S54384	envelope polyprote
21	2540.5	76.4	856	1 A44963	env polyprotein pr
22	2534.5	76.2	846	1 VCLJND	env polyprotein pr
23	2191	65.9	854	1 VCLJSI	env polyprotein pr
24	1969.5	59.2	506	2 A40218	envelop glycoprote
25	1945.5	58.5	445	2 A41621	env polyprotein M
26	1941.5	58.4	443	2 C41621	env polyprotein p
27	1849	55.6	454	2 B41621	env polyprotein D
28	1839	55.3	495	2 S31493	env polyprotein -
29	1803	54.2	877	2 S49197	envelope protein p

30	1719	51.7	863	2 A53034	gag polyprotein -
31	1301	39.1	290	2 S25940	env protein - huma
32	1286	38.7	297	2 S60538	envelope polyprote
33	1189	35.7	299	2 S60528	envelope polyprote
34	1188.5	35.7	294	2 S60545	envelope polyprote
35	1187.5	35.7	294	2 S60524	envelope polyprote
36	1186	35.6	299	2 S60529	envelope polyprote
37	1179	35.4	301	2 S60548	envelope polyprote
38	1173.5	35.3	300	2 S60546	envelope polyprote
39	1169.5	35.2	300	2 S60547	envelope polyprote
40	1169	35.1	299	2 S60521	envelope polyprote
41	1166.5	35.1	852	1 VCLJGG	env polyprotein pr
42	1166	35.0	859	1 VCLJST	env polyprotein pr
43	1165	35.0	877	2 C46356	env polyprotein -
44	1163	35.0	301	2 S60532	envelope polyprote
45	1163	35.0	301	2 S60531	envelope polyprote

ALIGNMENTS

RESULT 1
H44001
env polyprotein precursor - human immunodeficiency virus type 1 (strain YU-2)
N:Alternate names: coat polyprotein
N:Contains: coat protein gp120; coat protein gp41
C:Species: human immunodeficiency virus type 1, HIV-1
A>Note: host Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Sep-1994
C:Accession: H44001
R:Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.
J. Virol. 66, 6587-6600, 1992
A:Title: Complete nucleotide sequence, genome organization, and biological properties of
A:Reference number: A44001; MUID:93021387; PMID:1404605
A:Accession: H44001
A:Molecule type: DNA
A:Residues: 1-843 <LIY>
A:Cross-references: GB:M93258
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F:1-29/Domain: signal sequence #status predicted <SIG>
F:19-35/Region: hydrophobic
F:30-489/Product: coat protein gp120 #status predicted <GP1>
F:490-843/Product: coat protein gp41 #status predicted <GP2>
F:499-515/Region: hydrophobic
F:673-689/Region: hydrophobic
F:738-755/Domain: transmembrane #status predicted <TMN>
F:87,129,135,138,154,158,184,193,230,237,258,272,285,291,297,327,351,381,389,395,400,435,

Query Match 87.3%; Score 2906; DB 1; Length 843;
Best Local Similarity 84.2%; Pred. No. 1.4e-202;
Matches 544; Conservative 37; Mismatches 35; Indels 30; Gaps 5;

Qy	1	SAVEKLMWTVVYGVVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEI	VLEN 60
Db	28	SAAEQLMWTVVYGVVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEV	KLEN 87
Qy	61	VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSS	NWKEM 120
Db	88	VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNATNTSS	SWETM 147
Qy	121	DRGEIKNCSPKV-----GAGKLINCNTSVITQACPKVSFE	155
Db	148	EXGEIKNCSPNITTSIRDKVQKEYALFYNLDVVPIDNASYRLISCNTSVITQAC	PKVSFE 207
Qy	156	PIPIHYCAPAGFAILKCNDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLA	EEGVV 215
Db	208	PIPIHYCAPAGFAILKCNDKFKNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLA	EEIEV 267
Qy	216	IRSENFDTNAKTIIVQLKESVEINCTRPNNTNRKSIITIGPGRAFYATGDIIGD	IRQAHCN 275

Db 268 IRSENFNNAKTIIIVQLNESVVINCTRPNNNTRKINSIGPGRALYTTGEIIGDIRQAHCN 327

Qy 276 ISGEKWNNTLKQIVTKLQAQFG-NKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFN 334

Db 328 LSKTQWENTLEQIAIKLKEQFGNNKTIIFNPSSGGDPEIVTHSFNCGGEFFYCNSTQLP- 386

Qy 335 STWNTTIGPNNNTGITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDG 394

Db 387 -TWNDRKLNNTGRNITLPCRIKQIINMWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDG 445

Qy 395 GKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKAKRRVVQREKRAVTILGA 454

Db 446 GKD-TNGTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKAKRRVVQREKRAVGLGA 504

Qy 455 MFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQ 513

Db 505 LFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKOLQ 564

Qy 514 ARVLAVERYLKDQQLLGWCGSKLICTTAVPWNASWSNKSLDQIWNMTWMEWEREIDN 573

Db 565 ARVLAVERYLKDQQLLGWCGSKLICTTVPWNTSWSNKSLEIWDNMTWMKWEREIDN 624

Qy 574 YTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 625 YTHIIYSLIEQSONQQEKNEQELLALDKWASLWNWFDITKWLWYIK 670

RESULT 2

T09448

envelope glycoprotein - human immunodeficiency virus type 1 (strain JRFL)

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 24-Nov-1999

C;Accession: T09448

R;Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie, submitted to the EMBL Data Library, July 1996

A;Reference number: Z16673

A;Accession: T09448

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-847 <PAN>

A;Cross-references: EMBL:U63632; NID:g1465777; PID:g1465781

C;Genetics:

A;Gene: env

C;Superfamily: type E retrovirus env polyprotein

Query Match 87.3%; Score 2906; DB 2; Length 847;

Best Local Similarity 85.4%; Pred. No. 1.5e-202;

Matches 554; Conservative 25; Mismatches 38; Indels 32; Gaps 6;

Qy 1 SAVEKLVTVTVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60

Db 28 SAVEKLVTVTVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLEN 87

Qy 61 VTENFNWKNMWVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

Db 88 VTEHFNWKNMWVEQMQEDIISLWDQSLKPCVKLTPLCVTLNCKDV-NATNTNDSEGTM 146

Qy 121 DRGEIKNCSFKVAG-----KLINCNTSVITQACPVS 153

Db 147 ERGEIKNCSFNITTSIRDEVOKEYALFYKLDVVPIDNNTSYRLISCDTSVITQACPKIS 206

Qy 154 FEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEG 213

Db 207 FEPIPIHYCAPAGFAILKCNCKTFNKGCPKNVSTVQCTHGIRPVVSTQLLLNGSLAEEE 266

Qy 214 VVIRSENFNTDAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAYATGDIIGDIRQAH 273

Db 267 VVIRSDNFTNNAKTIIVQLKESVEINCTRPNNNTRKSIHIGPGRAYTTGEIIGDIRQAH 326

Qy 274 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF 333

Db 327 CNISRAKWNNTLKQIVIKLREQFENKTIVFNHSSGGDPEIVMHSFNCGGEFFYCNSTQLF 386

Qy 334 NSTW-NNTIGPNNNG-TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLT 391

Db 387 NSTWNNTEGSNNTEGNTITLPCRIKQIINMWQEVGKAMYAPPPIRGQIRCSSNITGLLLT 446

Qy 392 RDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKAKRRVVQREKRAVT 451

Db 447 RDGGIN-ENGTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKAKRRVVQREKRAVG 505

Qy 452 LGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK 511

Db 506 IGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQNNLLRAIEAQQRMQLQLTVWGIK 565

Qy 512 -LQARVLAVERYLKDQQLLGWCGSKLICTTAVPWNASWSNKSLDQIWNMTWMEWERE 570

Db 566 QLQARVLAVERYLGDQQLLGWCGSKLICTTAVPWNASWSNKSLDRIWNNMTWMEWERE 625

Qy 571 IDNYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 626 IDNYTSEIYTLIEESQNQQEKNEQELLELDKWASLWNWFDITKWLWYIK 674

RESULT 3

T12016

envelope glycoprotein - human immunodeficiency virus type 1 (strain sc14.3)

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C;Accession: T12016

R;McCutchan, F.E.; Sanders-Buell, E.; Salminen, M.O.; Carr, J.K.; Sheppard, W.H. AIDS Res. Hum. Retroviruses 14, 329-337, 1998

A;Title: Diversity of the human immunodeficiency virus type 1 envelope glycoprotein in S

A;Reference number: Z17379; MUID:98178716; PMID:9519894

A;Accession: T12016

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-852 <MCC>

A;Cross-references: EMBL:U90934; NID:g2351783; PIDN:AAC59271.1; PID:g2351784

C;Genetics:

A;Gene: env

C;Superfamily: type E retrovirus env polyprotein

Query Match 87.2%; Score 2900.5; DB 2; Length 852;

Best Local Similarity 84.3%; Pred. No. 3.7e-202;

Matches 552; Conservative 32; Mismatches 32; Indels 39; Gaps 7;

Qy 1 SAVEKLVTVTVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVTPDPNPQEIIVLEN 60

Db 28 SAAEQLRVTVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVTPDPNPQEVVLKN 87

Qy 61 VTENFNWKNMWVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTN-LKNATNTKSSNWKE 119

Db 88 VTENFNWKNMWVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDYLNDTNTTSNNGG 147

Qy 120 MDRGEIKNCSFKV-----GAGKLINCNTSVITQACPKV 152

Db 148 MEGGEIKNCSFNITTRIGNKVQKEYALFYKLDVVPIDNTTTSYRLINCNTSVITQACPKV 207

Qy 153 SFEPPIPIHYCAPAGFAILKCNCKKFNKSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEE 212

Db 208 SFEPPIHYCTPAGFALLKCKDKKFNKGTGCTNVSTVQCTHGIRPVVSTQLLLNGSLAEE 267

Qy 213 GVVIRSENFNTDAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAYATGDIIGDIRQA 272

Db 268 EVVIRSENFNTDAKTIIVQLNESVEINCTRPNNNTRKSIHIGPQALYATGAIIGDIRQA 327

Qy 273 HCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQL 332

Db 328 HCNISRAKWNNTLKQIVKLVQVFGNKTIIFNQSSGGDPEIVMHSFNCGGEFFYCNTTKL 387

Qy 333 FNSTW--NNTIGPNNNT-----NGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNI 385

Db 388 FNSTWMFNNTW--NDTETDENGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNI 445

Qy 386 TGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKAKRRVVQR 445

Db	446	TGILLTRDGGTN-NSTNETFRPGGDMRDNRSELYKYKVVKIEPLGVAPTRAKRRVVQR	504
Qy	446	EKRAVTLGAMFLGLFLGAAGSTMGARSLTLTVQARQLLSGIVQQNLLRAIEAQOHLQL	505
Db	505	EKRAVGIGALFLGLGAAGSTMGAAAVTLTVQARQLLSGIVQQNLLRAIEAQOHLQL	564
Qy	506	TVWGIK-LQARVLAVERYLKDQLLGIWCSGKLICTTAVPWNASWSKSLDQIWNMTW	564
Db	565	TVWGIKQLQARVLAVERYLRDQLLGIWCSGKLICTTTPWNASWSKSLDKIWNMTW	624
Qy	565	MEWERIDNYTNLIYTLIEESQNOQKNEQELLELDKWASLWNWFDISKWLWYIK	619
Db	625	MEWERIDNYTSLIYSLIEESQNOQKNEQELLELDKWASLWSWFDITKWLWYIK	679
RESULT 4			
S13289			
env protein - human immunodeficiency virus type 1			
C;Species: human immunodeficiency virus type 1, HIV-1			
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997			
C;Accession: S13289			
R;O'Brien, W.A.; Koyanagi, Y.; Namazie, A.; Zhao, J.Q.; Diagne, A.; Idler, K.; Zack, J.A			
Nature 348, 69-73, 1990			
A;Title: HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120			
A;Reference number: S13288; MUID:91043044; PMID:2172833			
A;Accession: S13289			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-847 <OBR>			
C;Superfamily: type E retrovirus env polyprotein			
Query Match	86.8%;	Score 2889;	DB 2; Length 847;
Best Local Similarity	84.9%;	Pred. No. 2.5e-201;	
Matches	551;	Conservative	24; Mismatches 42; Indels 32; Gaps 6;
Qy	1	SAVEKLVVTVYGYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN	60
Db	28	SATEKLVVTVYGYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLEN	87
Qy	61	VTENFNMWKNMVQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM	120
Db	88	VTENFNMWKNMVQMQEDIISLWDQSLKPCVKLTPLCVTLNCKDV-NATNTTNDSEGMT	146
Qy	121	DRGEIKNCSFKVGAG-----KLINCNTSVITQACPKVS	153
Db	147	ERGEIKNCSFNITTSIRDEVQKEYALFYKLDVVPIDNNNTSVRLISCDTSVITQACPKIS	206
Qy	154	FEPIPIHYCAPAGFAILKCNCKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLABEG	213
Db	207	FEPIPIHYCAPAGFAILKCNCKTFNGKGPCKNVSTVDCTHGIRPVVSTQLLNGSLABEE	266
Qy	214	VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH	273
Db	267	VVIRSDNFTNNAKTIIVQLKESVEINCTRPNNNTRKSIHIGPGRAYTTGEIIGDIRQAH	326
Qy	274	CNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF	333
Db	327	CNISRAKWNNTLKQIVIKLREQFNKTIIVFNHSSGGDPEIVMHSFNCGGEFFYCNSTQLN	386
Qy	334	NSTW-NNTIGPNTNG-TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLT	391
Db	387	NSTWNNTEGSNNTEGNTITLPCRIKQFINMWQEVGKAMYAPPPIRGQIRCSSNITGLLLT	446
Qy	392	RDGGEKISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVT	451
Db	447	RDGGIN-ENGTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVG	505
Qy	452	LGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQNLLRAIEAQOHLQLTVWGIK	511
Db	506	IGAFLGLGLGAAGSTMGAASMTLTVQARLLLSGIVQQNLLRAIEAQORMLQLTVWGIK	565
Qy	512	-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSKSLDQIWNMTWMEWERE	570

Db	566	QLQARVLAVERYLGDQQLLGIWCSGKLICTTAVPWNASWSKSLDRIWNNMTWMEWERE	625
Qy	571	IDNYTNLIYTLIEESQNOQKNEQELLELDKWASLWNWFDISKWLWYIK	619
Db	626	IDNYTSEIYTLIEESQNOQKNEQELLELDKWASLWNWFDITKWLWYIK	674
RESULT 5			
VCLJA2			
env polyprotein precursor - human immunodeficiency virus type 1 (isolate ARV-2)			
N;Alternate names: coat polyprotein			
C;Species: human immunodeficiency virus type 1, HIV-1			
A;Note: host Homo sapiens (man)			
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 16-Jul-1999			
C;Accession: A03976			
R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh			
Science 227, 484-492, 1985			
A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).			
A;Reference number: A04003; MUID:85090453; PMID:2578227			
A;Accession: A03976			
A;Molecule type: DNA			
A;Residues: 1-855 <SAN>			
A;Cross-references: GB:K02007; NID:g328658; PIDN:AAB59882.1; PID:g328666			
C;Genetics:			
A;Gene: env			
C;Superfamily: type E retrovirus env polyprotein			
C;Keywords: AIDS; capsid protein; coat protein; glycoprotein; immunodeficiency; polyprot			
F;1-30/Domain: signal sequence #status predicted <SIG>			
F;31-509/Product: exterior membrane glycoprotein #status predicted <EXT>			
F;510-855/Product: transmembrane glycoprotein #status predicted <TMM>			
F;87,129,140,158,184,190,200,244,265,292,298,304,334,341,358,364,388,394,400,408,445,458,			
F;610,624,636,815/Binding site: carbohydrate (Asn) (covalent) #status predicted			
Query Match	85.8%;	Score 2856;	DB 1; Length 855;
Best Local Similarity	82.3%;	Pred. No. 6.2e-199;	
Matches	540;	Conservative	32; Mismatches 46; Indels 38; Gaps 6;
Qy	1	SAVEKLVVTVYGYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN	60
Db	28	SATEKLVVTVYGYGVPVWKEATTTLFCASDARAYDTEVHNVWATHACVPTDPNPQEVVLGN	87
Qy	61	VTENFNMWKNMVQEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM	120
Db	88	VTENFNMWKNMVQEQEDIISLWDQSLKPCVKLTPLCVTLNCTDLGKATNTSSNWKEE	147
Qy	121	DRGEIKNCSFKVGAG-----KLINCNTSVITQA	148
Db	148	IKGEIKNCSFNITTSIRDKIOKENALFRNLDVVPIDNASTTNTYRLIHCNRSVITQA	207
Qy	149	CPKVSFEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNVSTVQCTHGIRPVVSTQLLNGS	208
Db	208	CPKVSFEPIPIHYCTPAGFAILKCNCKTFNGKGPCTNVSTVQCTHGIRPIVSTQLLNGS	267
Qy	209	LABEGVVIRSENFNDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAYATGDIIGD	268
Db	268	LABEEVIRSDNFTNNAKTIIVQLNESVAINCTRPNNNTRKSIYIGPGRAYHTTGRIIGD	327
Qy	269	IRQAHCNISGEKWNNTLKQIVTKLQAFG-NKTIIVFKQSSGGDPEIVMHSFNCGGEFFYC	327
Db	328	IRKAHCNISRAQWNNNTLEQIVKKLREQFGNNKTIIVFNQSSGGDPEIVMHSFNCRGEFFYC	387
Qy	328	NSTQLFNSTW--NNTIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNI	385
Db	388	NTQLFNNTWRLNHTEGTKG-NDTIILPCRIKQIINMWQEVGKAMYAPPIGGQISCSSNI	446
Qy	386	TGILLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQR	445
Db	447	TGILLTRDGGTNTVNTDTEVFRPGGDMRDNRSELYKYKVVKIEPLGIAPTAKRRVVQR	506
Qy	446	EKRAV-TLGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQNLLRAIEAQOHLLO	504
Db	507	EKRAVGIVGAMFLGLGAAGSTMGAVSLTLTVQARQLLSGIVQQNLLRAIEAQOHLLO	566

Db 448 IRCSSNITGLLLTRDGGNN--NNGSEIFRPPGGDMRDNRSELYKYKVVKIEPLGVAPTKA 506

Qy 439 KRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEA 498

Db 507 KRRVVQREKRAVGIGALFLGFLGAAGSTMGARSMTLTVQARQLLSGIVQQNNLLRAIEA 566

Qy 499 QOHLQLTWGIK-LQARVLAVERYLKDQQLGIWCGSGKLICTTAVPWNASWSNKSLDQ 557

Db 567 QOHLQLTWGIKQLOARILAVERYLKDQQLGIWCGSGKLICTTAVPWNASWSNKSLEQ 626

Qy 558 IWNNMTWMEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWY 617

Db 627 IWNNMTWMEWDREINNYTSLIHSLIEESQOQEKNEQELLELDKWASLWNWFNITNWLWY 686

Qy 618 IK 619

Db 687 IK 688

RESULT 8

S33985

env polyprotein - human immunodeficiency virus type 1

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999

C;Accession: S33985

R;Carlini, F.

submitted to the EMBL Data Library, November 1991

A;Reference number: S33979

A;Accession: S33985

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-851 <CAR>

A;Cross-references: EMBL:Z11530; NID:g60192; PIDN:CAA77628.1; PID:g60199

C;Superfamily: type E retrovirus env polyprotein

Query Match 85.3%; Score 2839.5; DB 2; Length 851;

Best Local Similarity 83.3%; Pred. No. 9.7e-198;

Matches 544; Conservative 32; Mismatches 40; Indels 37; Gaps 9;

Qy 1 SAVEKLVVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60

Db 29 SATEKLVVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEVVLVN 88

Qy 61 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKE- 119

Db 89 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMI 148

Qy 120 MDRGEIKNCSFKVGAGK-----LINCNTSVITQACPKV 152

Db 149 MEKGEIKNCSFNISTSKRGKVQKEYAFYKLDIIPIDNDTTSYLTSCNTSVITQACPKV 208

Qy 153 SFEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNVSTVQCTHGIRPVWSTQLLNGSLAE 212

Db 209 SFEPIPIHYCAPAGFAILKCNKNTFNGTGPCTNVSTVQCTHGIRPVWSTQLLNGSLAE 268

Qy 213 GVVIRSENFTDNAKTIIVOLKESVEINCTRPNNNTRKSITI--GPGRAFYATGDIIGDIR 270

Db 269 EVVIRSNFTDNAKTIIVQLNTSVEINCTRPNNNTRKKIRIQRGPGRAFTIGK-IGNMR 327

Qy 271 QAHCNISGEKWNNTLKQIVTKLQAFG-NKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNS 329

Db 328 QAHCNISRAKWNNTLKQIDSKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNS 387

Qy 330 TQLFNSTWNNTIGPNNTNG--TITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITG 387

Db 388 TQLFNSTW-STKGSNNTSGSDTITLPCRICKQIINMWQEVGKAMYAPPISGQIRCSSNITG 446

Qy 388 LLLTRDGGKEISNTTEIFRPPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREK 447

Db 447 LLLTRDGGNS--NNESEIFRPPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREK 505

Qy 448 RAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTV 507

Db 506 RAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTV 565

Qy 508 WGIK-LQARVLAVERYLKDQQLGIWCGSGKLICTTAVPWNASWSNKSLSLDQIWNNTWME 566

Db 566 WGIKQLOARILAVERYLKDQQLGIWCGSGKPICTTAVPWNASWSNKSLEQIWNNTWME 625

Qy 567 WEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 626 WDREINNYTSLIHSLIEESQOQEKNEQELLELDKWASLWNWFNITNWLWYIK 678

RESULT 9

VCLJVL

env polyprotein precursor - human immunodeficiency virus type 1 (isolate LV)

N;Alternate names: coat polyprotein

C;Species: human immunodeficiency virus type 1, HIV-1

A;Note: host Homo sapiens (man)

C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 16-Jul-1999

C;Accession: A03974

R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.

Nature 313, 450-458, 1985

A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi

A;Reference number: A93355; MUID:85111157; PMID:2982104

A;Accession: A03974

A;Molecule type: DNA

A;Residues: 1-856 <MUE>

A;Cross-references: GB:K02083; NID:g555008; PIDN:AAB59873.1; PID:g328559

C;Genetics:

A;Gene: env

C;Superfamily: type E retrovirus env polyprotein

C;Keywords: AIDS; capsid protein; coat protein; glycoprotein; immunodeficiency; polyprot

F;1-30/Domain: signal sequence #status predicted <SIG>

F;31-511/Product: exterior membrane glycoprotein #status predicted <EXT>

F;512-856/Product: transmembrane glycoprotein #status predicted <TMM>

F;88,136,141,156,160,186,197,230,234,241,262,276,289,295,301,332,339,356,386,392,397,406,

F;611,616,625,637,674,750,816/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 85.3%; Score 2838; DB 1; Length 856;

Best Local Similarity 83.0%; Pred. No. 1.3e-197;

Matches 545; Conservative 33; Mismatches 39; Indels 40; Gaps 10;

Qy 1 SAVEKLVVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60

Db 29 SATEKLVVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEVVLVN 88

Qy 61 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKE- 119

Db 89 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMI 148

Qy 120 MDRGEIKNCSFKVGA---GK-----LINCNTSVITQACPKV 152

Db 149 MEKGEIKNCSFNISTSRGKVQKEYAFYKLDIIPIDNDTTSYLTSCNTSVITQACPKV 208

Qy 153 SFEPIPIHYCAPAGFAILKCNCKKFNKSGPCTNVSTVQCTHGIRPVWSTQLLNGSLAE 212

Db 209 SFEPIPIHYCAPAGFAILKCNKNTFNGTGPCTNVSTVQCTHGIRPVWSTQLLNGSLAE 268

Qy 213 GVVIRSENFTDNAKTIIVOLKESVEINCTRPNNNTRKSITI--GPGRAFYATGDIIGDIR 270

Db 269 EVVIRSANFTDNAKTIIVQLNQSVEINCTRPNNNTRKSIRIQRGPGRAFTIGK-IGNMR 327

Qy 271 QAHCNISGEKWNNTLKQIVTKLQAFG-NKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNS 329

Db 328 QAHCNISRAKWNNTLKQIDSKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNS 387

Qy 330 TQLFNSTWNN---TIGPNNTNG--TITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSS 383

Db 388 TQLFNSTWFNSTWSTEGSNNTSGSDTITLPCRICKQFINMWQEVGKAMYAPPISGQIRCSS 447

Qy 384 NITGILLTRDGGKEISNTTEIFRPPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV 443

Db 448 NITGILLTRDGGNN--NNESEIFRPPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV 506

QY	444	QREKRAVTLGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHL	503
Db	507	QREKRAVGIGALFLGLGAAGSTMGAASMTLTVQARQLLSGIVQQNNLLRAIEAQOHL	566
QY	504	QLTVWGK-LQARVLAVERYLKDQLLGIWCSGKLICTTAVPWNASWSNKSLDQIWNM	562
Db	567	QLTVWGKQLQARILAVERYLKDQLLGIWCSGKLICTTAVPWNASWSNKSLEQIWNM	626
QY	563	TWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYK	619
Db	627	TWMEWDREINNYTSLIHSLIEESNQOEKNEQELLELDKWANLWNWLNITNWLWYK	683
RESULT 10			
VCLJSC			
env polyprotein precursor - human immunodeficiency virus type 1 (isolate SC)			
N;Alternate names: coat polyprotein			
C;Species: human immunodeficiency virus type 1, HIV-1			
A;Note: host Homo sapiens (man)			
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Feb-1997			
C;Accession: B28922			
R;Gurgo, C.; Guo, H.G.; Franchini, G.; Aldovini, A.; Collalti, E.; Farrell, K.; Wong-Sta			
Virology 164, 531-536, 1988			
A;Title: Envelope sequences of two new United States HIV-1 isolates.			
A;Reference number: A28922; MUID:88219542; PMID:3369091			
A;Accession: B28922			
A;Molecule type: DNA			
A;Residues: 1-861 <GUR>			
C;Genetics:			
A;Gene: env			
C;Superfamily: type E retrovirus env polyprotein			
C;Keywords: capsid protein; coat protein; glycoprotein; polyprotein; transmembrane prote			
F;1-29/Domain: signal sequence #status predicted <SIG>			
F;30-861/Product: env polyprotein #status predicted <EPP>			
F;87,129,135,140,143,159,163,187,198,234,241,262,276,295,301,302,333,340,356,362,386,396			
Query Match 85.2%; Score 2836; DB 1; Length 861;			
Best Local Similarity 82.3%; Pred. No. 1.8e-197;			
Matches 545; Conservative 30; Mismatches 43; Indels 44; Gaps 9;			
QY	1	SAVEKLVTVVYGYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIVLEN	60
Db	28	SAAEQLVTVVYGYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLGN	87
QY	61	VTFENFMWKNMVQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLK-----NATNTKSS	115
Db	88	VTFENFMWKNMVQMHEDIISLWDSLKPCVKLTPLCVTLNCTNLRNDTSTNATNTSS	147
QY	116	NWKEMDRGEIKNCSEFKVGAGK-----LINCNTSVITQACP	150
Db	148	NRGKMEGGEWNTCSFNTTSRSKVQKEYALFYKLDVVPIDNDTTSYTLINCNTSVITQACP	207
QY	151	KVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLA	210
Db	208	KVSFEPIPIHYCA-RWFAILNCNNKKFNGTGPCNTNVSTVQCTHGIRPVVSTHLLNGSLA	266
QY	211	EEGVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSITI--GPGRAFYATGDIIGD	268
Db	267	EEEVVLRSENFDTNAKTIIVQLKEAVEINCTRPNNNTTRSIHIOQGPGRAFYATGDIIGD	326
QY	269	IRQAHCNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCN	328
Db	327	IRQAHCNISRAKWNNTLKQIVIKLRDQFENKTIIFNRSSGGDPEIVMHSFNCGGEFFYCN	386
QY	329	STQLFNSWNTNNTIGP---NNTNG--TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS	383
Db	387	STQLFSSTWNGTEGSTKGNNTGGNDTITLFCRIKEIINMWQEVGKAMYAPPPIKGQVKCSS	446
QY	384	NITGLLLTRDGGKEISNT---TEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAK	439
Db	447	NITGLLLTRDGGNSKNGSKNENTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAK	506
QY	440	RRVVQREKRAV-TLGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEA	498

Db	507	RRVVQREKRAVGTIGAMFLGLGAAGSTMGATSMTLTVQARLLLSGIVQQNNLLRAIEA	566
QY	499	QQHLLQLTVWGK-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLDQ	557
Db	567	QQHLLQLTVWGKQLQARVLAVERYLRDQQLLGIWCSGKLICTTVPWNTSWSNKS�KD	626
QY	558	IWNNTWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWY	617
Db	627	IWGNMTWMEWEREIDNYTSLIYTLIEESNQOEKNEQELLELDKWASLWNWFNITNWLWY	686
QY	618	IK 619	
Db	687	IK 688	
RESULT 11			
S13288			
env protein - human immunodeficiency virus type 1			
C;Species: human immunodeficiency virus type 1, HIV-1			
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997			
C;Accession: S13288			
R;O'Brien, W.A.; Koyanagi, Y.; Namazie, A.; Zhao, J.Q.; Diagne, A.; Idler, K.; Zack, J.A			
Nature 348, 69-73, 1990			
A;Title: HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120			
A;Reference number: S13288; MUID:91043044; PMID:2172833			
A;Accession: S13288			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-854 <OBR>			
C;Superfamily: type E retrovirus env polyprotein			
Query Match 85.1%; Score 2830; DB 2; Length 854;			
Best Local Similarity 82.7%; Pred. No. 4.8e-197;			
Matches 542; Conservative 32; Mismatches 43; Indels 38; Gaps 9;			
QY	1	SAVEKLVTVVYGYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIVLEN	60
Db	29	SATEKLVTVVYGYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLVN	88
QY	61	VTFENFMWKNMVQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKVNATNTKSSNWKE-	119
Db	89	VTFENFMWKNMDMVEQMHEDIISLWDSLKPCVKLTPLCVSLKCTDLKNDTNTNSSGRMI	148
QY	120	MDRGEIKNCSEFKVGAG-----KLINCNTSVITQACPKVSF	154
Db	149	MEKEIKNCSEFNISTSIRDVKQKEYAFYKLDIVPIDNTSYRLISCNTSVITQACPKVSF	208
QY	155	EPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGV	214
Db	209	EPIPIHYCAPAGFAILKCNCKNTFNGTGPCNTNVSTVDCTHGIRPVVSTQLLNGSLAEEVD	268
QY	215	VIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSITI--GPGRAFYATGDIIGDIRQA	272
Db	269	VIRSANFTDNAKTIIVQLNTSVVEINCTRPNNNTRKSIRIQRGPGRAFVTIGK-IGNMRQA	327
QY	273	HCNISGEKWNNTLKQIVTKLQAFG-NKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQ	331
Db	328	HCNISRAKWNATLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQ	387
QY	332	LFNSTWNN----TIGPNNTNG--TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNI	385
Db	388	LFNSTWFNSTWSTEGSNTEGSDTITLPCRIKQFINMWQEVGKAMYAPPISGQIRCSSNI	447
QY	386	TGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQR	445
Db	448	TGLLLTRDGGNN-NNGSEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQR	506
QY	446	EKRAVTLGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHLLOL	505
Db	507	EKRAVGIGALFLGLGAAGSTMGCTSMTLTVQARQLLSDIVQQNNLLRAIEAQOHLLOL	566
QY	506	TVWGIK-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLDQIWNMTW	564

Db 567 TVWGIKQLQARILAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLEQIWNMTW 626

Qy 565 MEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 627 MEWDREINNYTSLIHSLIEESQOQEKNEQELLELDKWASLWNWFNITNWLWYIK 681

RESULT 12

VCLJJB

env polyprotein - human immunodeficiency virus type 1 (isolate BR)

N:Alternate names: coat polyprotein

N:Contains: coat protein gp120; coat protein gp41

C:Species: human immunodeficiency virus type 1, HIV-1

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Feb-1997

C:Accession: A31667

R:Anand, R.; Thayer, R.; Srinivasan, A.; Nayyar, S.; Gardner, M.; Luciw, P.; Dandekar, S

Virology 168, 79-89, 1989

A:Title: Biological and molecular characterization of human immunodeficiency virus (HIV-

A:Reference number: A94389; MUID:89085613; PMID:2789516

A:Accession: A31667

A:Molecule type: DNA

A:Residues: 1-852 <ANA>

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: capsid protein; coat protein; polyprotein; transmembrane protein

F:1-516/Product: coat protein gp120 #status predicted <CP1>

F:517-852/Product: coat protein gp41 #status predicted <CP2>

Query Match 84.7%; Score 2817; DB 1; Length 852;

Best Local Similarity 81.6%; Pred. No. 4.1e-196;

Matches 533; Conservative 38; Mismatches 46; Indels 36; Gaps 7;

Qy 1 SAVEKLMVTVYYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60

Db 29 SATDKLMVTVYYGVPVWKEANTTLFCASDAKAYDTEIHNWATHACVPTDPNPQELVMGN 88

Qy 61 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

Db 89 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCLTLNCHDF-NATNATSNSGKWM 147

Qy 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACP 151

Db 148 EGGEMKNCSEFNITTSIRDKMQKEYALFYKLDIVPIDNDKNTNRYLISCNTSVITQACP 207

Qy 152 VSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAE 211

Db 208 VTFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAE 267

Qy 212 EGVVIRSENFTDNAKTIIIVQLKESVEINCTRPNNTRKSIITGPGRAFYATGDIIGDIRQ 271

Db 268 EEVIRSENFTNNVKTIIIVQLNESVEINCTRPNNTRKRITMGPRVYYTTGQIIGDIRR 327

Qy 272 AHCNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIIVMHSFNCGEFFYCNSTQ 331

Db 328 AHCNLSRSKWENTLKQIVTKLRVQFNKTIIVFNRSSGGDPEIIVMHSFNCGEFFFCNTTQ 387

Qy 332 LFNSTW-NNTIGPNNTNGT--ITLPCRKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGL 388

Db 388 LFNSTWYRNTTG-NITEGNSPITLPCRKQIINMWQEVGKAMYAPPPIRGQIKCSSNITGL 446

Qy 389 LLTRDGGKEISNT-TEIFRPGGGMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREK 447

Db 447 LLTRDGGNNNETTDETEIFRPGGGMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREK 506

Qy 448 RAVTLGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTV 507

Db 507 RAVGLGALFLGLGAAGSTMGAASLTTLTVQARLLSGIVQQNNLLMAIEAQHMLELTV 566

Qy 508 WGIK-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLSLDQIWNMTWME 566

Db 567 WGIKQLQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLSLDIWDNMTWME 626

Qy 567 WEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 627 WEREIDNYTNLIYSLIEDSQIOQEKNEKELLELDKWASLWNWFNITNWLWYIK 679

RESULT 13

VCLJH4

env polyprotein - human immunodeficiency virus type 1 (isolate CDC-451)

N:Alternate names: coat polyprotein

N:Contains: coat protein gp120; coat protein gp41

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999

C:Accession: C25523

R:Desai, S.M.; Kalyanaraman, V.S.; Casey, J.M.; Srinivasan, A.; Andersen, P.R.; Devare, S

Proc. Natl. Acad. Sci. U.S.A. 83, 8380-8384, 1986

A:Title: Molecular cloning and primary nucleotide sequence analysis of a distinct human

A:Reference number: A94136; MUID:87041461; PMID:3490666

A:Accession: C25523

A:Molecule type: DNA

A:Residues: 1-868 <DES>

A:Cross-references: GB:M13137; NID:g326460; PIDN:AAA44311.1; PID:g326467

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: capsid protein; coat protein; glycoprotein; polyprotein; transmembrane protei

F:1-521/Product: coat protein gp120 #status predicted <GP1>

F:522-868/Product: coat protein gp41 #status predicted <GP2>

F:89,138,139,142,166,195,198,208,245,252,273,287,300,312,342,349,365,371,395,405,409,459

Query Match 84.6%; Score 2813.5; DB 1; Length 868;

Best Local Similarity 79.3%; Pred. No. 7.6e-196;

Matches 533; Conservative 36; Mismatches 44; Indels 59; Gaps 8;

Qy 1 SAVEKLMVTVYYGVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60

Db 30 SAAANLWVTVYYGVPVWKEATTLFCASDAKAYDTEAHNVWATHACVPTNPNPQEVLEN 89

Qy 61 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCT--NLKNATNTKSSN-- 116

Db 90 VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLNCTDLNNTNTTNTLSII 149

Qy 117 --WKEMDRGEIKNCSFKVGAG-----KLINCNT 142

Db 150 VWEOQKGMKNCSEFNITTSIRDKVQREYALFYKLDVEPIDDNKNTNTTKYRLINCNT 209

Qy 143 SVITQACPKVSEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQ 202

Db 210 SVITQACPKVSEPIPIHYCTPTGFALLKCNCKKFGNGTGPCTNVSTVQCTHGIRPVVSTQ 269

Qy 203 LLLNGSLAEVIRSENFTDNAKTIIIVQLKESVEINCTRPNNTRKSIITGPGRAFYAT 262

Db 270 LLLNGSLAEVIRSENFTNNAKTIIIVQLNVSVEINCTRPNNHTRKRVTLGPRVWYTT 329

Qy 263 GDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIIVMHSFNCGG 322

Db 330 GEILGNIRQAHCNISRAQWNNTLQQIATTLREQFGNKTIAFNQSSGGDPEIIVMHSFNCGG 389

Qy 323 EFFYCNSTQLFNSTWNNTIGPNNTNGT-----ITLPCRKQIINRWQEVGKA 369

Db 390 EFFYCNSTQLFNSAWNVT-----SNGTWSVTRKQKDTGDIITLPCRKQIINRWQVVGKA 444

Qy 370 MYAPPPIRGQIRCSSNITGLLTRDGGKEISNTTEIFRPGGGMRDNRSELYKYKVVKIE 429

Db 445 MYALPIKGLIRCSSNITGLLTRDGGGE-NQTTEIFRPGGGMRDNRSELYKYKVVKIE 503

Qy 430 PLGVAPTAKRRVVQREKRAV-TLGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQ 488

Db 504 PLGVAPTAKRRVVQREKRAVGMGLGAMFLGLGAAGSTMGATSMALTQVARQLLSGIVQQ 563

Qy 489 QNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWN 547

Db 564 QNNLLRAIKAQHLLQLTVWGIKQLQARILAVERYLKDQQLLGFWCSGKLICTTAVPWN 623

Qy	548	ASWSNKSILDQIWNMTWMEWEREIDNYTNLIYTLIEESONQOEKNEQELLELDKWASLWN	607
Db	624	ASWSNKTLDQIWNMTWMEWDREIDNYTHLIYTLIEESONQOEKKNQELLQLDKWASLWT	683
Qy	608	WFDISKWLWYIK	619
Db	684	WSDITKWLWYIK	695
RESULT 14			
VCLJJK			
env polyprotein precursor - human immunodeficiency virus type 1 (strain KB-1-gp32)			
N;Alternate names: coat polyprotein			
N;Contains: coat protein gp120; coat protein gp32			
C;Species: human immunodeficiency virus type 1, HIV-1			
A;Note: host Homo sapiens (man)			
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 01-Mar-1996			
C;Accession: B42995			
R;Shimizu, H.; Hasebe, F.; Tsuchie, H.; Morikawa, S.; Ushijima, H.; Kitamura, T.			
Virology 189, 534-546, 1992			
A;Title: Analysis of a human immunodeficiency virus type 1 isolate carrying a truncated			
A;Reference number: A42995; MUID:92351552; PMID:1322587			
A;Accession: B42995			
A;Molecule type: mRNA			
A;Residues: 1-729 <SHI>			
A;Cross-references: GB:S41266; GB:D01206			
C;Genetics:			
A;Gene: env			
C;Superfamily: type E retrovirus env polyprotein			
C;Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein			
F;1-689/Domain: extracellular #status predicted <EXT>			
F;1-33/Domain: signal sequence #status predicted <SIG>			
F;17-33/Region: hydrophobic #status predicted			
F;34-517/Product: coat protein gp120 #status predicted <CP1>			
F;514-517/Region: cleavage processing #status predicted			
F;518-729/Product: coat protein gp32 #status predicted <CP2>			
F;518-534/Region: hydrophobic #status predicted			
F;690-711/Domain: transmembrane #status predicted <TM1>			
F;712-729/Domain: intracellular #status predicted <INT>			
F;93,141,145,146,163,191,192,237,241,248,269,283,308,338,345,361,367,397,403,408,414			
Query Match 83.9%; Score 2792.5; DB 1; Length 729;			
Best Local Similarity 79.3%; Pred. No. 2e-194;			
Matches 525; Conservative 41; Mismatches 47; Indels 49; Gaps 7;			
Qy	1	SAVEKLWTVVYGVVPVWKEATTLFLCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN	60
Db	34	SAAEQLVTVVYGVVPVWKEATTLFLCASDAKAYDTEAHNVWATHACVPTDPNPQEVVLN	93
Qy	61	VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM	120
Db	94	VTENFNWKNMVEQMHENIISLWDQSLKPCVKLTPLCVTLHCTDLRNTTNNSSIEEKM	153
Qy	121	DRGEIKNCSFKVGAG-----KLINCNTSVITQAC	149
Db	154	-KGEIKNCSFNVTNIRDVKQKEYALFYKLDVVPIDNDNNSTNTCYRLISCDTSVITQAC	212
Qy	150	PKVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSL	209
Db	213	PKVSFEPIPIHYCTPAGFALLKCNKNTFNGTGPKCNVSTVQCTHGIRPVVSTQLLLNGSL	272
Qy	210	AEEGVVIRSEFTDNAKTIIVOLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDI	269
Db	273	AEEGVVIRSEFTDNVKTIIIVQLNETVKINCIRPNNKTRKRVTMGPRVYTTGEIIGDI	332
Qy	270	ROAHCNISGEKNWNTLKQIVTKLQAFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNS	329
Db	333	ROAHCNISRAEWNKLTLEQIANLKRQFENKTIIVFNQSSGGDPEIVMHNFCGGEFFYCD	392
Qy	330	TQLFNST-----WNNTIGPNNTNGTITLPCRICKQIINRWQEVGKAMYAPPIRGQIRCS	382
Db	393	SQLFNSTHLSNGTWWNGT-GPEN----ITLPCRICKQIIVNMWQEVGKAMYAPPIRGQIRCS	447

Qy	383	SNITGLLLTRDGGKEISNNT-----EIFRPGGGMDMRDNWRSELYKYKVVKIEPLGVAPT	KA 438
Db	448	SNITGLLLTRDGGNTQNNNTNSSIEIFRPGGGMDMRDNWRSELYKYKVVKIEPLGVAP	TRA 507
Qy	439	KRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIE	A 498
Db	508	KRRVVQREKRAVGIGAVFLGFLGAAGSTMGAAAVTLTVQARQLLPGIVQQNNLLRAID	A 567
Qy	499	QOHLQLTVWGIK-LQARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKS	LDQ 557
Db	568	QOHLQLTVWGIKQLQARVLAVERYLKDQQLMGWCSGKFICTTAVPWNTSWSNKS	FPNE 627
Qy	558	IWNMTWMEWEREIDNYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWFDISK	WLWY 617
Db	628	IWDNMTWMEWEREINNNTNLIYNLIEESONQOEKNEQDLLALDKWDSLWNWFSIT	TKWLWY 687
Qy	618	IK 619	
Db	688	IK 689	
RESULT 15			
VCLJKB			
env polyprotein precursor - human immunodeficiency virus type 1 (strain KB-1-gp41)			
N;Alternate names: coat polyprotein			
N;Contains: coat protein gp120; coat protein gp41			
C;Species: human immunodeficiency virus type 1, HIV-1			
A;Note: host Homo sapiens (man)			
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 01-Mar-1996			
C;Accession: A42995			
R;Shimizu, H.; Hasebe, F.; Tsuchie, H.; Morikawa, S.; Ushijima, H.; Kitamura, T.			
Virology 189, 534-546, 1992			
A;Title: Analysis of a human immunodeficiency virus type 1 isolate carrying a truncated			
A;Reference number: A42995; MUID:92351552; PMID:1322587			
A;Accession: A42995			
A;Molecule type: mRNA			
A;Residues: 1-861 <SHI>			
A;Cross-references: GB:S41266; GB:D01206			
C;Genetics:			
A;Gene: env			
C;Superfamily: type E retrovirus env polyprotein			
C;Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein			
F;1-689/Domain: extracellular #status predicted <EXT>			
F;1-33/Domain: signal sequence #status predicted <SIG>			
F;17-33/Region: hydrophobic #status predicted			
F;34-517/Product: coat protein gp120 #status predicted <CP1>			
F;514-517/Region: cleavage processing #status predicted			
F;518-861/Product: coat protein gp41 #status predicted <CP2>			
F;518-534/Region: hydrophobic #status predicted			
F;690-711/Domain: transmembrane #status predicted <TM1>			
F;712-861/Domain: intracellular #status predicted <INT>			
F;756-772/Region: hydrophobic #status predicted			
F;93,141,145,146,163,191,192,237,241,248,269,283,296,308,338,345,361,367,397,403,408,414,			
Query Match 83.9%; Score 2792.5; DB 1; Length 861;			
Best Local Similarity 79.3%; Pred. No. 2.5e-194;			
Matches 525; Conservative 41; Mismatches 47; Indels 49; Gaps 7;			
Qy	1	SAVEKLWTVVYGVVPVWKEATTLFLCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN	60
Db	34	SAAEQLVTVVYGVVPVWKEATTLFLCASDAKAYDTEAHNVWATHACVPTDPNPQEVVLN	93
Qy	61	VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM	120
Db	94	VTENFNWKNMVEQMHENIISLWDQSLKPCVKLTPLCVTLHCTDLRNTTNNSSIEEKM	153
Qy	121	DRGEIKNCSFKVGAG-----KLINCNTSVITQAC	149
Db	154	-KGEIKNCSFNVTNIRDVKQKEYALFYKLDVVPIDNDNNSTNTCYRLISCDTSVITQAC	212
Qy	150	PKVSFEPIPIHYCAPAGFAILKCNCKKFGSGPCTNVSTVQCTHGIRPVVSTQOLLNGSL	209
Db	213	PKVSFEPIPIHYCTPAGFALLKCNKNTFNGTGPKCNVSTVQCTHGIRPVVSTQOLLNGSL	272

Qy	210	AEEGVIRSENFTD	NAKTIIVQLKESVEIN	TRPNNNTRKSITIGP	GRAFYATGDIIGDI	269
Db	273	AEEGVIRSENFTD	NKTIIVQLNETVKINC	IRPNNKTRKRVMTG	PGRVYTTGEIIGDI	332
Qy	270	ROAHCNISGEKWN	NTLKQIVTKLOAQFG	NKTIIVFKOSSGGDP	EIVMHSFNCGGEFFYCNS	329
Db	333	ROAHCNISRAEW	NKTLEQIANLKRKQF	ENKTIIVFNQSSGGDP	EIVMHNFNCGGEFFYCDS	392
Qy	330	TQLFNST-----	WNNTIGPNNNTNGT	ITLPCRIKQIINRWQ	EVGKAMYAPPPIRGQIRCS	382
Db	393	SQLFNSTHLSNGT	WNGT-GPEN-----	ITLPCRIKQIIVNMW	QEVGKAMYAPPPIRGQIRCS	447
Qy	383	SNITGLLLTRDGG	KEISNTT-----	EIFRPGGDMRDNR	SELYKYKVVKIEPLGVAPTKA	438
Db	448	SNITGLLLTRDGG	NTQNNTNSSIEIFR	PGGDMRDNRSELY	KYKVVKIEPLGVAPTRA	507
Qy	439	KRRVVQREKRAV	TGLGAMFLGLGAAG	STMGARSLTLTVQAR	QLLSGIVQQNNLLRAIEA	498
Db	508	KRRVVQREKRAV	GIGAVFLGLGAAG	STMGAAAVTLTVQAR	QLLPGIVQQNNLLRAIDA	567
Qy	499	QOHLQLTVWG	IK-LOARVLAVERY	LKDQQLLGWCSGK	LICTTAVPWNASWSNKS	LDQ 557
Db	568	QOHLQLTVWG	IKQLQARVLAVERY	LKDQQLMGWCSGK	FICTTAVPWNTSWSNKS	PFNE 627
Qy	558	IWNNTWMEW	EREIDNYTNLIYTL	IEESQNQEKNEQEL	LIEDKWLWFDISKWLWY	617
Db	628	IWDNMTWME	WEREINNYTNLIY	NLIEESQNQEKNEQ	DLALDKWDSLWNF	SITKWLWY 687
Qy	618	IK	619			
Db	688	IK	689			

This Page Blank (uspto)

FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	274	274	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	293	293	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	299	299	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	329	329	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	336	336	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	388	388	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	398	398	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	401	401	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	438	438	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	454	454	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	602	602	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	607	607	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	628	628	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
SQ	SEQUENCE	847 AA;	96135 MW;	0A901317FD7FF2AB	CRC64;

Query Match	98.4%;	Score 3275;	DB 1;	Length 847;
Best Local Similarity	95.2%;	Pred. No. 3.5e-246;		
Matches 616;	Conservative 0;	Mismatches 3;	Indels 28;	Gaps 2;

Qy	1	SAVEKLVVTYYGVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEI	VLEN 60
Db <td>28</td> <td>SAVEKLVVTYYGVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEI<td>VLEN 87</td></td>	28	SAVEKLVVTYYGVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEI <td>VLEN 87</td>	VLEN 87
Qy <td>61</td> <td>VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNA<td>TNKSNNWEM 120</td></td>	61	VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNA <td>TNKSNNWEM 120</td>	TNKSNNWEM 120
Db <td>88</td> <td>VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNA<td>TNKSNNWEM 147</td></td>	88	VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNA <td>TNKSNNWEM 147</td>	TNKSNNWEM 147
Qy <td>121</td> <td>DRGEIKNCSFKVGAG-----KLN<td>CNTSVITQACPKVS 153</td></td>	121	DRGEIKNCSFKVGAG-----KLN <td>CNTSVITQACPKVS 153</td>	CNTSVITQACPKVS 153
Db <td>148</td> <td>DRGEIKNCSFKVTTIRNKMKEYALFYKLDVVPIDNDNTSYKL<td>INCNTSVITQACPKVS 207</td></td>	148	DRGEIKNCSFKVTTIRNKMKEYALFYKLDVVPIDNDNTSYKL <td>INCNTSVITQACPKVS 207</td>	INCNTSVITQACPKVS 207
Qy <td>154</td> <td>FEPIPIHYCAPAGFAILKNDKKFNGSGPCTNVSTVQCTHGIRPVWSTQ<td>LLNGSLAEEG 213</td></td>	154	FEPIPIHYCAPAGFAILKNDKKFNGSGPCTNVSTVQCTHGIRPVWSTQ <td>LLNGSLAEEG 213</td>	LLNGSLAEEG 213
Db <td>208</td> <td>FEPIPIHYCAPAGFAILKNDKKFNGSGPCTNVSTVQCTHGIRPVWSTQ<td>LLNGSLAEEG 267</td></td>	208	FEPIPIHYCAPAGFAILKNDKKFNGSGPCTNVSTVQCTHGIRPVWSTQ <td>LLNGSLAEEG 267</td>	LLNGSLAEEG 267
Qy <td>214</td> <td>VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITGPGRAFYATG<td>DIIGDIRQAH 273</td></td>	214	VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITGPGRAFYATG <td>DIIGDIRQAH 273</td>	DIIGDIRQAH 273
Db <td>268</td> <td>VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITGPGRAFYATG<td>DIIGDIRQAH 327</td></td>	268	VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITGPGRAFYATG <td>DIIGDIRQAH 327</td>	DIIGDIRQAH 327
Qy <td>274</td> <td>CNISGEKNNTLKQIVTKLQAQFGNKTIIVFKQSSGGDPEIVMHSFNC<td>GGEFFYCNSTQLF 333</td></td>	274	CNISGEKNNTLKQIVTKLQAQFGNKTIIVFKQSSGGDPEIVMHSFNC <td>GGEFFYCNSTQLF 333</td>	GGEFFYCNSTQLF 333
Db <td>328</td> <td>CNISGEKNNTLKQIVTKLQAQFGNKTIIVFKQSSGGDPEIVMHSFNC<td>GGEFFYCNSTQLF 387</td></td>	328	CNISGEKNNTLKQIVTKLQAQFGNKTIIVFKQSSGGDPEIVMHSFNC <td>GGEFFYCNSTQLF 387</td>	GGEFFYCNSTQLF 387
Qy <td>334</td> <td>NSTWNNTIGPNNNTGTTILPCRIKQIINRWQEVGKAMYAPPIRGQIR<td>CSSNITGLLLTRD 393</td></td>	334	NSTWNNTIGPNNNTGTTILPCRIKQIINRWQEVGKAMYAPPIRGQIR <td>CSSNITGLLLTRD 393</td>	CSSNITGLLLTRD 393
Db <td>388</td> <td>NSTWNNTIGPNNNTGTTILPCRIKQIINRWQEVGKAMYAPPIRGQIR<td>CSSNITGLLLTRD 447</td></td>	388	NSTWNNTIGPNNNTGTTILPCRIKQIINRWQEVGKAMYAPPIRGQIR <td>CSSNITGLLLTRD 447</td>	CSSNITGLLLTRD 447
Qy <td>394</td> <td>GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRV<td>VQREKRAVTLG 453</td></td>	394	GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRV <td>VQREKRAVTLG 453</td>	VQREKRAVTLG 453
Db <td>448</td> <td>GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRV<td>VQREKRAVTLG 507</td></td>	448	GGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRV <td>VQREKRAVTLG 507</td>	VQREKRAVTLG 507
Qy <td>454</td> <td>AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQQNNLLRAIEAQH<td>LLQLTVWGIK-L 512</td></td>	454	AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQQNNLLRAIEAQH <td>LLQLTVWGIK-L 512</td>	LLQLTVWGIK-L 512
Db <td>508</td> <td>AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQQNNLLRAIEAQH<td>LLQLTVWGIKQL 567</td></td>	508	AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQQNNLLRAIEAQH <td>LLQLTVWGIKQL 567</td>	LLQLTVWGIKQL 567
Qy <td>513</td> <td>QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSLLDQI<td>WNNMTMEWEREID 572</td></td>	513	QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSLLDQI <td>WNNMTMEWEREID 572</td>	WNNMTMEWEREID 572
Db <td>568</td> <td>QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSLLDQI<td>WNNMTMEWEREID 627</td></td>	568	QARVLAVERYLKDQQLLGWCSGKLICTTAVPWNASWSNKSLLDQI <td>WNNMTMEWEREID 627</td>	WNNMTMEWEREID 627
Qy <td>573</td> <td>NYTNLIYTLIEESQNQEKNEQELLELDKWASLWNNWFDISKWLWYIK</td> <td>619</td>	573	NYTNLIYTLIEESQNQEKNEQELLELDKWASLWNNWFDISKWLWYIK	619
Db <td>628</td> <td>NYTNLIYTLIEESQNQEKNEQELLELDKWASLWNNWFDISKWLWYIK</td> <td>674</td>	628	NYTNLIYTLIEESQNQEKNEQELLELDKWASLWNNWFDISKWLWYIK	674

AC	P35961;
DT	01-JUN-1994 (Rel. 29, Created)
DT	01-JUN-1994 (Rel. 29, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN	ENV.
OS	Human immunodeficiency virus type 1 (YU-2 isolate) (HIV-1).
OC	Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX	NCBI_TaxID=36377;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=93021387; PubMed=1404605;
RA	Li Y., Hui H., Burgess C.J., Price R.W., Sharp P.M., Hahn B.H., Shaw G.M.;
RA	"Complete nucleotide sequence, genome organization, and biological properties of human immunodeficiency virus type 1 in vivo: evidence for limited defectiveness and complementation.";
RT	J. Virol. 66:6587-6600(1992).
CC	- - - - -
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch). -----
DR	EMBL; M93258; -, NOT_ANNOTATED_CDS.
DR	PIR; H44001; H44001.
DR	PDB; 1G9N; 27-DEC-00.
DR	InterPro; IPR000328; Env GP41.
DR	InterPro; IPR000777; GP120.
DR	Pfam; PF00516; GP120; 1.
DR	Pfam; PF00517; GP41; 1.
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal; 3D-structure.
KW	SIGNAL
FT	CHAIN 1 29 CHAIN 30 489 EXTERIOR MEMBRANE GLYCOPROTEIN. CHAIN 490 843 TRANSMEMBRANE GLYCOPROTEIN. TRANSMEM 738 755 POTENTIAL. DISULFID 53 73 BY SIMILARITY. DISULFID 118 201 BY SIMILARITY. DISULFID 125 192 BY SIMILARITY. DISULFID 130 155 BY SIMILARITY. DISULFID 214 243 BY SIMILARITY. DISULFID 224 235 BY SIMILARITY. DISULFID 292 326 BY SIMILARITY. DISULFID 373 432 BY SIMILARITY. DISULFID 380 405 BY SIMILARITY. CARBOHYD 87 87 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 129 129 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 154 154 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 158 158 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 193 193 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 230 230 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 258 258 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 291 291 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 389 389 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 395 395 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 435 435 N-LINKED (GLCNAC. .) (POTENTIAL). CARBOHYD 450 450 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 598 598 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 843 AA; 95648 MW; C69DFD971C918B71 CRC64;

Query Match 87.3%; Score 2906; DB 1; Length 843;
Best Local Similarity 84.2%; Pred. No. 1.5e-217;
Matches 544; Conservative 37; Mismatches 35; Indels 30; Gaps 5;

Qy 1 SAVEKLVWTVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPOEIVLEN 60
Db 28 SAAEQLVWTVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPOEVKLEN 87

Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
Db 88 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNATNTSSSWETM 147

Qy 121 DRGEIKNCSFKV-----GAGKLINCNTSVITQACPKVSFE 155
Db 148 ERGEIKNCSFNITTSIRDKVQKEYALFYNDVVPIDNASYRLISCNTSVITQACPKVSFE 207

Qy 156 PIPHYCAPAGFAILKCNDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEGVV 215
Db 208 PIPHYCAPAGFAILKCNDKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEIV 267

Qy 216 IRSENFNTNAKTIIVQLKESVEINCTRPNNTRKSIITGPGRAFYATGDIIGDIRQAHCN 275
Db 268 IRSENFNTNAKTIIVQLNESVVINCTRPNNTRKSIINIGPRALYTTGEIIGDIRQAHCN 327

Qy 276 ISGEKWNNTLKQIVTKLQAOFG-NKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFN 334
Db 328 LSKTQWENTLEQIAIKLKEQFGNKKTIIFNPSSGGDPEIVTHSFNCGGEFFYCNSTQLF- 386

Qy 335 STWNNTIGPNNNTGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDG 394
Db 387 -TWNDTRKLNNTGRNITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRDG 445

Qy 395 GKEISNTTEIFRPGGDMRDNRWSELYKYKVVKIEPLGVAPTAKRRVWQREKRAVTLGA 454
Db 446 GKD-TNGTEIFRPGGDMRDNRWSELYKYKVVKIEPLGVAPTAKRRVWQREKRAVGLGA 504

Qy 455 MFLGFLGAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-LQ 513
Db 505 LFLGFLGAGSTMGAASITLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQLQ 564

Qy 514 ARVLAVERILKQQLLGIWCGSKLICCTTAVPWNASWSNKSLDQIWNMTWMEWEREIDN 573
Db 565 ARVLAVERILRQQLLGIWCGSKLICCTTVPWNTSWSNKSLEIWDNMTWMKWEREIDN 624

Qy 574 YTNLIYTLIEESNQEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 625 YTHIIVSLIEQSQNQEKNEQELLALDKWASLWNWFDITKWLWYIK 670

RESULT 3
ENV_HV1SC
ID ENV_HV1SC STANDARD; PRT; 856 AA.
AC P05878;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (SC isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88219542; PubMed=3369091;
RA Gurgo C., Guo H.-G., Franchini G., Aldovini A., Collalti E.,

RA Farrell K., Wong-Staal F., Gallo R.C., Reitz M.S. Jr.;
"Envelope sequences of two new United States HIV-1 isolates.";
RL Virology 164:531-536(1988).
CC -!- MISCELLANEOUS: THE SC ISOLATE WAS TAKEN FROM AN ARC PATIENT IN
CC 1984 IN SOUTHERN CALIFORNIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17450; -; NOT_ANNOTATED_CDS.
DR HIV; M17450; ENV\$SC.
DR InterPro; IPR000328; Env_GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW Signal.
FT SIGNAL 1 29
FT CHAIN 30 510 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 511 856 TRANSMEMBRANE GLYCOPROTEIN.
FT SITE 760 760 IN-FRAME TERMINATION CODON.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 206 BY SIMILARITY.
FT DISULFID 125 197 BY SIMILARITY.
FT DISULFID 130 160 BY SIMILARITY.
FT DISULFID 219 247 BY SIMILARITY.
FT DISULFID 228 239 BY SIMILARITY.
FT DISULFID 296 330 BY SIMILARITY.
FT DISULFID 376 439 BY SIMILARITY.
FT DISULFID 383 412 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 674 674 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 856 AA; 97055 MW; DAF4DA600EBA7A08 CRC64;

Query Match 86.6%; Score 2880.5; DB 1; Length 856;
Best Local Similarity 83.3%; Pred. No. 1.5e-215;
Matches 547; Conservative 31; Mismatches 40; Indels 39; Gaps 7;


```
Qy 1 SAVEKLWTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60
Db 28 SAAEQWLWTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEVVLGN 87

Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLK-----NATNTKSS 115
Db 88 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTNLNDTSTNATNTTSS 147

Qy 116 NWKEMDRGEIKNCSFKVGAG-----KLINCNTSVITQACP 150
Db 148 NRGKMEGGEMTNCSEFNITTSIRSKVQKEYALFYKLDVVPIDNTSYTLINCNTSVITQACP 207

Qy 151 KVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPCTNVSTVQCTHGIRPVVSTQLLLNGSLA 210
Db 208 KVSFEPIPIHYCA-RWFAILNCNNKKFNCGTGPCTNVSTVQCTHGIRPVVSTHLLNGSLA 266

Qy 211 EEGVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSITIGPGRAFAYATGDIIGDIR 270
Db 267 EEVVLRSSENFDTNAKTIIVQLKEAVEINCTRPNNNTTRSIHIGPGRAFAYATGDIIGDIR 326

Qy 271 QAHCNISGEKNWNTLKQIVTKLQAQFGNKTIIVFKOSSGGDPEIVMHSFNCGGEFFYCNS 330
Db 327 QAHCNISRAKNWNTLKQIVIKLRDQFENKTIIFNRSGGDPEIVMHSFNCGGEFFYCNS 386

Qy 331 QLFNSTWNNTIGPNNTNG--TITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGL 388
Db 387 QLFSSTWNGTEGSNNTGGNDTITLPCRICKBIINMWQEVGKAMYAPPPIKGQVKCSSNITGL 446

Qy 389 LLTRDGGKEISNT----TEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQ 444
Db 447 LLTRDGGNSKNGSKNENTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQ 506

Qy 445 REKRAV-TLGAMFLGFLGAAGSTMGARSILTVQARQLLSGIVQQNNLLRAIEAQHLL 503
Db 507 REKRAVGTIGAMFLGFLGAAGSTMGATSMTLTVQARLLSGIVQQNNLLRAIEAQHLL 566

Qy 504 QLTWVGIK-LQARVLAVERYLKDQQLLGITWCGSGKLICTTAVPNWNASWSNKSLDQIWNM 562
Db 567 QLTWVGIKQLQARVLAVERYLRDQQLLGITWCGSGKLICTTTVPWNTSWSNKSLDKIWNM 626

Qy 563 TMWEWEREIDVNTLIYTLIEESQKQKNEQELLELDKWLWYIK 619
Db 627 TMWEWEREIDVNTSLIYTLIEESQKQKNEQELLELDKWLWYIK 683

RESULT 4
ENV_HV1JR
ID _ENV_HV1JR STANDARD; PRT; 848 AA.
AC P20871;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (JRCSF isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11688;
RN [1]
RP SEQUENCE FROM N.A.
RA Koyanagi S., Chen I.S.Y.;
RL Submitted (DEC-1988) to the HIV data bank.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38429; AAB03749.1; -.
DR PDB; 1CE4; 18-MAR-99.
```

```
DR HIV; M38429; ENV$JRCSF.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal;
KW 3D-structure.
FT SIGNAL 1 32 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 33 503 TRANSMEMBRANE GLYCOPROTEIN.
FT CHAIN 504 848
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 203 BY SIMILARITY.
FT DISULFID 125 194 BY SIMILARITY.
FT DISULFID 130 154 BY SIMILARITY.
FT DISULFID 216 245 BY SIMILARITY.
FT DISULFID 226 237 BY SIMILARITY.
FT DISULFID 294 328 BY SIMILARITY.
FT DISULFID 374 437 BY SIMILARITY.
FT DISULFID 381 410 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 137 137 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 185 185 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 195 195 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 274 274 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 440 440 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 608 608 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 808 808 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 848 AA; 96475 MW; 20767F51227EC3F3 CRC64;

Query Match 86.3%; Score 2870.5; DB 1; Length 848;
Best Local Similarity 83.7%; Pred. No. 8.6e-215;
Matches 544; Conservative 30; Mismatches 43; Indels 33; Gaps 5;

Qy 1 SAVEKLWTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60
Db 28 SAAEQWLWTVVYGVVPVWKEATTTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEVVLGN 87

Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKVNATNTKSSNWKEM 120
Db 88 VTDEFNMWKNMVEQMVEDVINLWDQSLKPCVKLTPLCVTLNCKDV-NATNTTSSEGMM 146

Qy 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACP 152
Db 147 ERGEIKNCSFNITKSIRDKVQKEYALFYKLDVVPIDNKNNTKYRLISCNTSVITQACP 206

Qy 153 SFEPIPIHYCAPAGFAILKCNCKKFNCGSPCTNVSTVQCTHGIRPVVSTQLLLNGSLAE 212
Db 207 SFEPIPIHYCAPAGFAILKCNCKKFNCGQCKNVSTVQCTHGIRPVVSTQLLLNGSLAE 266

Qy 213 GVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSITIGPGRAFAYATGDIIGDIRQA 272
Db 267 KVVIRSDNFDTNAKTIIVQLNESVKINCTRPSNNTKRKSIHIGPGRAFYTGTGEIIGDIRQA 326
```



```
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 507 QREKRAVGIGALFLGFLGAAGSTMGAASMTLTQARQLLSGIVQQNLLRAIEAQHLL 566

Qy 504 QLTVMGK-LQARVLAVERYLKQQLLGWCGSGKLICTTAVPWNASWSNKSLSLDQIWNM 562
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 567 QLTVMGKQLQARILAVERYLKQQLLGWCGSGKLICTTAVPWNASWSNKSLEQIWNM 626

Qy 563 TTMWEAREIDNNTNLIYTLIEESQOQKNEQELLELDKWASLWNWFDISKWLWYIK 619
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 627 TTMWEWDREINNYSLSLHIEESQOQKNEQELLELDKWASLWNWFNITNWLWYIK 683

RESULT 7
ENV_HV1BR
ID ENV HV1BR STANDARD; PRT; 861 AA.
AC P03377;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (BRU isolate) (HIV-1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11686;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85099333; PubMed=2981635;
RA Wain-Hobson S., Sonigo P., Danos O., Cole S., Alizon M.;
RT "Nucleotide sequence of the AIDS virus, LAV.";
RL Cell 40:9-17(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02013; AAB59751.1; -.
DR EMBL; A04321; CAA00352.1; -.
DR PIR; A03975; VCLJLV.
DR PDB; 1ERF; 20-FEB-02.
DR HIV; K02013; ENV$BRU.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal;
KW 3D-structure.
FT SIGNAL 1 30
FT CHAIN 31 516 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 517 861 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 54 74 BY SIMILARITY.
FT DISULFID 119 210 BY SIMILARITY.
FT DISULFID 126 201 BY SIMILARITY.
FT DISULFID 131 162 BY SIMILARITY.
FT DISULFID 223 252 BY SIMILARITY.
FT DISULFID 233 244 BY SIMILARITY.
FT DISULFID 301 336 BY SIMILARITY.
FT DISULFID 383 450 BY SIMILARITY.
FT DISULFID 390 423 BY SIMILARITY.
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 642 642 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 821 821 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 861 AA; 97487 MW; 04DE2B4D4E4FD63A CRC64;

Query Match 85.6%; Score 2848.5; DB 1; Length 861;
Best Local Similarity 82.8%; Pred. No. 4.5e-213;
Matches 548; Conservative 30; Mismatches 39; Indels 45; Gaps 10;

Qy 1 SAVEKLVVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 29 SATEKLVVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLVN 88

Qy 61 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTKSSNWKE- 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 89 VTENFNMMKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLGNATNTSSNTNSS 148

Qy 120 -----MDRGEIKNCSPKVG---GK-----LINCNTSVITQ 147
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 149 SGEMMEKGEIKNCSEFNISTSRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVITQ 208

Qy 148 ACPKVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPCTNVSTVQCTHGIRPVVSTQLLNG 207
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 209 ACPKVSFEPIPIHYCAPAGFAILKCNCKTNGTGPCTNVSTVQCTHGIRPVVSTQLLNG 268

Qy 208 SLAEEGVVIRSEFTDNAKTIIVQLKESVEINCTRPNNNRKSIIT--GPGRAFYATGDI 265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 269 SLAEEVVIIRSANFTDNAKTIIVQLNQSVSEINCTRPNNNRKSIIRQGPGRFAVTIGK- 327

Qy 266 IGDIRQAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIVFKQSSGGDPEIVMHSFNCGGEP 324
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 328 IGNMQAHCNISRAKNATLKQIASKLREQFGNKTIIIFKQSSGGDPEIVTHSFNCGGEP 387

Qy 325 FYCNSITQLFNSTWNN-----TIGPNNTNG--TITLPCRICKQIINRWQEVGKAMYAPPISGQ 378
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 388 FYCNSITQLFNSTWFSNSTWSTEGSNNTGSDTITLPCRICKQFINMWQEVGKAMYAPPISGQ 447

Qy 379 IRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTKA 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 448 IRCSSNITGLLLTRDGGNN-NNGSEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTKA 506

Qy 439 KRRVVQREKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNLLRAIEA 498
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 507 KRRVVQREKRAVGIGALFLGFLGAAGSTMGARSMTLTQARQLLSGIVQQNLLRAIEA 566

Qy 499 QQHLLQLTVWGIK-LQARVLAVERYLKDQLLGWCGSGKLICTTAVPWNASWSNKSLSLQ 557
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 567 QQHLLQLTVWGIKQLQARILAVERYLKDQLLGWCGSGKLICTTAVPWNASWSNKSLSLQ 626

Qy 558 IWNMTWMEWEREIDNNTNLIYTLIEESQOQKNEQELLELDKWASLWNWFDISKWLWY 617
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 627 IWNMTWMEWDREINNYSLSLHIEESQOQKNEQELLELDKWASLWNWFNITNWLWY 686

Qy 618 IK 619
```


RX MEDLINE=87299196; PubMed=3040055;
RA Ratner L., Fisher A., Jagodzinski L.L., Mitsuya H., Liou R.-S.,
RA Gallo R.C., Wong-Staal F.;
RT "Complete nucleotide sequences of functional clones of the AIDS
virus.";
RL AIDS Res. Hum. Retroviruses 3:57-69(1987).
RN [2]
RP REVISIONS.
RA Ratner L., Fisher A., Jagodzinski L.L., Mitsuya H., Liou R.-S.,
RA Gallo R.C., Wong-Staal F.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K03455; AAB50262.1; -;
DR EMBL; AF038399; AAB99976.1; -;
DR EMBL; AF033819; AAC82596.1; -;
DR PDB; 1DF4; 26-JAN-00.
DR PDB; 1DF5; 26-JAN-00.
DR PDB; 1DLB; 02-APR-00.
DR PDB; 1G9M; 27-DEC-00.
DR PDB; 1GC1; 19-AUG-98.
DR PDB; 1GZL; 10-OCT-02.
DR PDB; 1K33; 10-OCT-01.
DR PDB; 1K34; 10-OCT-01.
DR HIV; K03455; ENV\$HXB2.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal;
3D-structure.
FT SIGNAL 1 30
FT CHAIN 31 511
FT CHAIN 512 856
FT CHAIN 54 74
FT DISULFID 119 205
FT DISULFID 126 196
FT DISULFID 131 157
FT DISULFID 218 247
FT DISULFID 228 239
FT DISULFID 296 331
FT DISULFID 378 445
FT DISULFID 385 418
FT CARBOHYD 88 88
FT CARBOHYD 136 136
FT CARBOHYD 141 141
FT CARBOHYD 156 156
FT CARBOHYD 160 160
FT CARBOHYD 186 186
FT CARBOHYD 197 197
FT CARBOHYD 230 230
FT CARBOHYD 234 234
FT CARBOHYD 241 241
FT CARBOHYD 262 262
FT CARBOHYD 276 276
FT CARBOHYD 289 289
FT CARBOHYD 295 295
FT CARBOHYD 301 301
FT CARBOHYD 332 332
FT CARBOHYD 339 339
FT CARBOHYD 356 356
FT CARBOHYD 386 386
FT CARBOHYD 392 392
FT CARBOHYD 397 397
FT CARBOHYD 406 406
FT CARBOHYD 448 448
FT CHAIN 512 856
FT CHAIN 54 74
FT DISULFID 119 205
FT DISULFID 126 196
FT DISULFID 131 157
FT DISULFID 218 247
FT DISULFID 228 239
FT DISULFID 296 331
FT DISULFID 378 445
FT DISULFID 385 418
FT CARBOHYD 88 88
FT CARBOHYD 136 136
FT CARBOHYD 141 141
FT CARBOHYD 156 156
FT CARBOHYD 160 160
FT CARBOHYD 186 186
FT CARBOHYD 197 197
FT CARBOHYD 230 230
FT CARBOHYD 234 234
FT CARBOHYD 241 241
FT CARBOHYD 262 262
FT CARBOHYD 276 276
FT CARBOHYD 289 289
FT CARBOHYD 295 295
FT CARBOHYD 301 301
FT CARBOHYD 332 332
FT CARBOHYD 339 339
FT CARBOHYD 356 356
FT CARBOHYD 386 386
FT CARBOHYD 392 392
FT CARBOHYD 397 397
FT CARBOHYD 406 406
FT CARBOHYD 448 448

FT CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 674 674 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 750 750 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 856 AA; 97212 MW; 6FAB16AF85107FE0 CRC64;
Query Match 85.2%; Score 2834; DB 1; Length 856;
Best Local Similarity 82.6%; Pred. No. 5.9e-212;
Matches 543; Conservative 33; Mismatches 41; Indels 40; Gaps 9;
QY 1 SAVEKLWTVVYGVVPMKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVLEN 60
Db |||||
QY 29 SATEKLWTVVYGVVPMKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLN 88
Db |||||
QY 61 VTENFMWKNMVEQMHEDIISLWQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKE- 119
Db |||||
QY 89 VTENFMWKNMVEQMHEDIISLWQSLKPCVKLTPLCVSLKCTDLKNDTNTSSGRMI 148
QY 120 MDRGEIKNCSFKVGAG-----KLNKNTSVITQACPKV 152
Db |||||
QY 149 MEKGEIKNCSFNISTIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKV 208
QY 153 SPEPIPIHYCAPAGFAILKCNKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEE 212
Db |||||
QY 209 SPEPIPIHYCAPAGFAILKCNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEE 268
QY 213 GVIRSENFTDNAKTIIVQLKESVEINCTRPNNTRKSIIT- -GPGRAFAYATGDIIGDIR 270
Db |||||
QY 269 EVVIRSVNFTDNAKTIIVQLNTSVEINCTRPNNTRKIRIQRGPGRAFVTIGK-IGNMR 327
QY 271 QAHCNISGEKWNNTLKQIVTKLQAQFG-NKTIIVFKQSSGDPPEIVMHSFNCGGEFFYCNS 329
Db |||||
QY 328 QAHCNISRAKNNTLKQIASKLREQFGNKKTIIFKQSSGDPPEIVTHSFNCGGEFFYCNS 387
QY 330 TQLFNSTWN- -TIGPNNTNG- -TITLPCRIKQIINRWQEVGKAMYAPPIRGQIRCSS 383
Db |||||
QY 388 TQLFNSTWTFNSTWSTEGSNNTSGSDTITLPCRIKQIINRWQKVGKAMYAPPIRGQIRCSS 447
QY 384 NITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYVVKIEPLGVAPTAKARRVV 443
Db |||||
QY 448 NITGLLLTRDGGNS- -NNESEIFRPGGDMRDNRSELYKYVVKIEPLGVAPTAKARRVV 506
QY 444 QREKRAVTLGAMFLGFLGAGSTMGARSLLTIVQARQLLSGIVQQNNLLRAIEAQHLL 503
Db |||||
QY 507 QREKRAVGIGALFLGFLGAGSTMGAASMTLTIVQARQLLSGIVQQNNLLRAIEAQHLL 566
QY 504 QLTWGIK-LQARVLAVERYLKDOQLLGIWCGSGKLICTTAVPWNASWSNKSLLDQIWNM 562
Db |||||
QY 567 QLTWGIKQLOARILAVERYLKDOQLLGIWCGSGKLICTTAVPWNASWSNKSLEQIWNHT 626
QY 563 TWMEWERIDNYTNLIYTLIEESQOQEKNEQELLELDKWLWYIK 619
Db |||||
QY 627 TWMEWDREINNYTSLIHSLEESQOQEKNEQELLELDKWLWYIK 683
RESULT 11
ENV_HV1B8
ID_HV1B8
AC P04582;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Envelope glycoprotein GP160 precursor [Contains: Exterior membrane
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (BH8 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11684;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=85111123; PubMed=2578615;
RA Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B.R.,
RA Josephs S.F., Doran E.R., Rafalski J.A., Whitehorn E.A.,
RA Baumeister K., Ivanoff L., Petteway S.R. Jr., Pearson M.L.,
RA Lautenberger J.A., Papas T.S., Ghayeb J., Chang N.T., Gallo R.C.,
RA Wong-Staal F.;
RT "Complete nucleotide sequence of the AIDS virus, HTLV-III.";
RL Nature 313:277-284(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02011; AAA44661.1; --
DR PDB; 1DDH; 13-JAN-99.
DR PDB; 1HHG; 31-OCT-93.
DR PDB; 1QO3; 02-JAN-00.
DR PDB; 1S2T; 24-DEC-97.
DR HIV; K02011; ENVSBH8.
DR GlycoSuiteDB; P04582; --
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal;
KW 3D-structure.
FT SIGNAL 1 30
FT CHAIN 31 506 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 507 851 TRANSMEMBRANE GLYCOPROTEIN.
FT DISULFID 54 74 BY SIMILARITY.
FT DISULFID 119 205 BY SIMILARITY.
FT DISULFID 126 196 BY SIMILARITY.
FT DISULFID 131 157 BY SIMILARITY.
FT DISULFID 218 247 BY SIMILARITY.
FT DISULFID 228 239 BY SIMILARITY.
FT DISULFID 296 331 BY SIMILARITY.
FT DISULFID 378 440 BY SIMILARITY.
FT DISULFID 385 413 BY SIMILARITY.
FT CARBOHYD 88 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 160 160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 632 632 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 851 AA; 96644 MW; D16A3C90857785F1 CRC64;

Query Match 85.0%; Score 2829.5; DB 1; Length 851;
Best Local Similarity 83.0%; Pred. No. 1.3e-211;
Matches 542; Conservative 33; Mismatches 41; Indels 37; Gaps 9;
QY 1 SAVEKLWVTYYVYGVVWKEATTLFCASDAKAYDTEVHNWVATHACVPTDPNPQEVLEN 60
DB 29 SATEKLWVTYYVYGVVWKEATTLFCASDAKAYDTEVHNWVATHACVPTDPNPQEVLEN 88
QY 61 VTENFNWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKE- 119
DB 89 VTENFNWKNMVEQMHEDIISLWDSLKPCVKLTPLCVSLKCTDLKNDTNTNSSGRMI 148
QY 120 MDRGEIKNCSFKVGAGK-----LINCNTSVITQACPKV 152
DB 149 MEKGEIKNCSFNISTSKRGKVQKEYAFFYKLDIIPIDNDTTSYLTSCNTSVITQACPKV 208
QY 153 SFEPIHYCAPAGFAILKCNKPFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 212
DB 209 SFEPIHYCAPAGFAILKCNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 268
QY 213 GVVIRSENFTDNAKTIIVQLKESVEINCTRPNNTRKSIIT--GPGRAFYATGDIIGDIR 270
DB 269 EVVIRSVNFTDNAKTIIVQLDTSVEINCTRPNNTRKKIRIQRGPGRAFTIGK-IGNMR 327
QY 271 QAHCNISGEKWNNTLKQIVTKLQAFG-NKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNS 329
DB 328 QAHCNISRAKNATLKQIDSKLREQFGNNKTIIFKQSSGGDPEIVHSFNCGGEFFYCNS 387
QY 330 TOLFNSTWNNTIGPNTNG--TITLPCRKQIINRWQEVGKAMVAPPVIRGQIRCSSNITG 387
DB 388 TOLFNSTW-STKGSNTEGSDTITLPCRKQIINRWQEVGKAMVAPPVIRGQIRCSSNITG 446
QY 388 LLLTRDGGKBSNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQREK 447
DB 447 LLLTRDGGNS-NNSEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQREK 505
QY 448 RAVTLGAMFLGLGAAGSTMARSLLTVQARQLLSGIVQQQNNLLRAIEAQHLLQLTV 507
DB 506 RAVGIGALFLGLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQHLLQLTV 565
QY 508 WGIK-LQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLDQIWNNTWME 566
DB 566 WGIKQLQARILAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLEQIWNNTWME 625
QY 567 WEREIDNYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWFDISKWLWYIK 619
DB 626 WDREINNYTSLIHSLEESONQOEKNEQELLELDKWASLWNWFTNWLWYIK 678
RESULT 12
ENV_HV1BN STANDARD; PRT; 852 AA.
ID ENV_HV1BN
AC P12488;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Envelope polypeptide GP160 precursor [Contains: Exterior membrane
DE glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (1BR isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11693;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89085613; PubMed=2789516;
RA Anand R., Thayer R., Srinivasan A., Nayyar S., Gardner M., Luciw P.,
RA Dandekar S.;
RT "Biological and molecular characterization of human immunodeficiency
RT virus (HIV-1BR) from the brain of a patient with progressive
RT dementia.";
RL Virology 168:79-89(1989).
CC -!- MISCELLANEOUS: THIS VIRUS IS CYTOPATHICALLY ACTIVE AND WAS

DR	InterPro; IPR000777; GP120.	1	29	
DR	Pfam; PF00516; GP120; 1.	30	509	
DR	Pfam; PF00517; GP41; 1.	510	855	
KW	AIDS; Coat protein; Polypeptide; Glycoprotein; Transmembrane; Signal.			
KW	SIGNAL	1	29	
FT	CHAIN	30	509	EXTERIOR MEMBRANE GLYCOPROTEIN.
FT	CHAIN	510	855	TRANSMEMBRANE GLYCOPROTEIN.
FT	DISULFID	53	73	BY SIMILARITY.
FT	DISULFID	118	210	BY SIMILARITY.
FT	DISULFID	125	201	BY SIMILARITY.
FT	DISULFID	130	162	BY SIMILARITY.
FT	DISULFID	223	223	BY SIMILARITY.
FT	DISULFID	233	244	BY SIMILARITY.
FT	DISULFID	301	335	BY SIMILARITY.
FT	DISULFID	381	442	BY SIMILARITY.
FT	DISULFID	388	415	BY SIMILARITY.
FT	CARBOHYD	87	87	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	134	134	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	142	142	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	161	161	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	165	165	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	192	192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	202	202	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	246	246	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	267	267	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	281	281	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	300	300	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	306	306	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	336	336	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	359	359	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	389	389	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	395	395	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	399	399	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	405	405	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	458	458	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	610	610	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	624	624	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	636	636	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	815	815	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	855 AA; 97476 MW; 9CF82A607ADD62DA CRC64;		
Query Match 84.7%; Score 2818; DB 1; Length 855;				
Best Local Similarity 80.3%; Pred. No. 1e-210;				
Matches 526; Conservative 41; Mismatches 52; Indels 36; Gaps 4;				
QY	1	SAVEKLWTVVYGVVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVLEN	60	
DB	28	SAENLWTVVYGVVWKEATTLFCASDARAYATEVHNWATHACVPTDPNPQEVVLGN	87	
QY	61	VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCT-----NLKNATNTK	113	
DB	88	VTENFDMWKNMVEQMEDIISLWDQSLKPCVKLTPLCVTLCTDVTSSSLRNATNTT	147	
QY	114	SSNWKEMDRGEIKNCSFKVGAG-----KLINCNTSVIT	146	
DB	148	SSSWETMEKELKNCSFNTTTSIRDKMQEQYALFYKLDVLPIDKNDTKFRLIHCNTSTIT	207	
QY	147	QACPKEVFEPIPIHYCAPAGFAILKNCNKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLN	206	
DB	208	QACPKEVFEPIPIHYCTPAGFAILKNCNKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLN	267	
QY	207	GSLAEEGVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDII	266	
DB	268	GSLAEEVEIIRSNFTNNAKIIIVQLNKSVEINCTRPNNNTRNRIISIGPGRAFHTTKQII	327	
QY	267	GDIROAHNCNISEKWNNTLKQIVTKLQAOFGNKTIVFKQSSGGDPEIVMHSFNCGGGEFFY	326	
DB	328	GDIROAHNCNLSRATWEKTLQIATKLKQFNRKNTIADRSSGGDPEIVMHSFNCGGGEFFY	387	

QY	327	CNSTQFNSTWNTIGPNNTNGTITLPCRIKQIINRWQEVGKAMYPPIRGQIRCSSNIT	386	
DB	388	CNTSOLFSTWNTIGPNNTNGTITLPCRIKQIINRWQEVGKAMYPPIRGQIRCSSKIT	447	
QY	387	GLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQRE	446	
DB	448	GLLLTRDGGKNTTNGIEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQRE	507	
QY	447	KRAV-TLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQL	505	
DB	508	KRAVGLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQL	567	
QY	506	TVWGK-LQARVLAVERYLKDQQLLGWCGSKLICCTTAVPWNASWSNKSLSQIWNMTW	564	
DB	568	TVWGKQLQARVLAVERYLKDQQLLGWCGSKLICCTTVPWNASWSNKSLSLNEIWDNMTW	627	
QY	565	MEWEREIDNTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK	619	
DB	628	QWQEREIDNTNLIYTLIEESQNQQEKNEQELLELDKWAGLWSWFSITNWLWYIR	682	
RESULT 14				
ENV_HVIC4	ID	ENV_HVIC4	STANDARD;	PRT; 868 AA.
AC	P05879;			
DT	01-NOV-1988 (Rel. 09, Created)			
DT	01-NOV-1988 (Rel. 09, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].			
GN	ENV.			
OS	Human immunodeficiency virus type 1 (CDC-451 isolate) (HIV-1).			
OC	Viruses; Retrovirus; Retroviridae; Lentivirus.			
OX	NCBI_TaxID=11687;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87041461; PubMed=3490666;			
RA	Desai S.M., Kalyanaram V.S., Casey J.M., Srinivasan A., Andersen P.R., Devare S.G.;			
RT	"Molecular cloning and primary nucleotide sequence analysis of a distinct human immunodeficiency virus isolate reveal significant divergence in its genomic sequences."			
RT	Proc. Natl. Acad. Sci. U.S.A. 83:8380-8384 (1986).			
RL	[2]			
RN	SEQUENCE OF 34-43.			
RP	MEDLINE=90253924; PubMed=2187500;			
RX	Kalyanaram V.S., Rodriguez V., Veronese F., Rahman R., Lusso P., DeVico A.L., Copeland T., Oroszlan S., Gallo R.C., Sarngadharan M.G.;			
RT	"Characterization of the secreted, native gp120 and gp160 of the human immunodeficiency virus type 1."			
RL	AIDS Res. Hum. Retroviruses 6:371-380 (1990).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL; M13137; AAA44311.1; --			
DR	PIR; C25523; VCLJH4.			
DR	HIV; M13137; ENVSCDC45.			
DR	InterPro; IPR000328; Env GP41.			
DR	InterPro; IPR000777; GP120.			
DR	Pfam; PF00516; GP120; 1.			
DR	Pfam; PF00517; GP41; 1.			
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal.			
KW	SIGNAL	1	33	
FT	CHAIN	34	522	EXTERIOR MEMBRANE GLYCOPROTEIN (GP120).
FT	CHAIN	523	868	TRANSMEMBRANE GLYCOPROTEIN.

FT	DISULFID	55	75	BY SIMILARITY.	
FT	DISULFID	120	216	BY SIMILARITY.	
FT	DISULFID	127	207	BY SIMILARITY.	
FT	DISULFID	132	163	BY SIMILARITY.	
FT	DISULFID	229	258	BY SIMILARITY.	
FT	DISULFID	239	250	BY SIMILARITY.	
FT	DISULFID	307	341	BY SIMILARITY.	
FT	DISULFID	387	456	BY SIMILARITY.	
FT	DISULFID	394	429	BY SIMILARITY.	
FT	CARBOHYD	89	89	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	131	131	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	139	139	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	142	142	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	162	162	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	166	166	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	195	195	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	198	198	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	208	208	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	245	245	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	252	252	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	273	273	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	287	287	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	300	300	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	306	306	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	312	312	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	342	342	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	349	349	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	365	365	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	371	371	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	395	395	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	405	405	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	409	409	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	459	459	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	628	628	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	637	637	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	649	649	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	828	828	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	868 AA;	98698 MW;	A11527FC52A6F0C8	CRC64;
Query Match 84.6%; Score 2813.5; DB 1; Length 868;					
Best Local Similarity 79.3%; Pred. No. 2.4e-210;					
Matches 533; Conservative 36; Mismatches 44; Indels 59; Gaps 8;					
QY	1	SAVEKLVVTVYYGVPVWKEATTTLCASDAKAYDTEVHNWATHACVPTDPNPQEIIVLEN	60		
Db	30	SAANLVTVYYGVPVWKEATTTLCASDAKAYDTEAHNVWATHACVTPNPNQEVVLEN	89		
QY	61	VTENFNWKNMVQEHEDIISLWDQSLKPCVKLTPLCVTLHCT--NLKNATNFKSSN--	116		
Db	90	VTENFNWKNMVQEHEDIISLWDQSLKPCVKLTPLCVTLNCTDLTNNTTNTTELSII	149		
QY	117	--WKEMDRGEIKNCSEFKVGAG-----KLINCNT	142		
Db	150	VVWEQRGKGEMRNCSEFNITTSIRDKVQREYALFYKLDVEPIDDNKNTTNTKVRINCNT	209		
QY	143	SVITQACPKVSFEPIPIHYCAPAGFAILKNCDDKFKNSGPGCTNVSTVQCTHGIRPVVSTQ	202		
Db	210	SVITQACPKVSFEPIPIHYCTPTGTGFALLKNCDDKFKNGTGPCTNVSTVQCTHGIRPVVSTQ	269		
QY	203	LLNGSLAEEGVIRSENFNTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYAT	262		
Db	270	LLNGSLAEEEVIRSENFNTNAKTIIVQLNVSVVEINCTRPNNHTRKRVTLGPGRVWYTT	329		
QY	263	GDIIGDIRQAHCNISGEKWNNTLKQIVTKLQAOFGNKTIIVEKQSSGGDPEIVMHSFNCGG	322		
Db	330	GEILGNIRQAHCNISRAQWNNTLQQIATTLREQFGNKTIAFNQSSGGDPEIVMHSFNCGG	389		
QY	323	EFFYCNSTQLFNSTWNNTIGPNNNTGT-----ITLPCRICKQIINRWQEVGKA	369		
Db	390	EFFYCNSTQLFNSAANNVT-----SNGTWSVTRKQKDTGDIITLPCRICKQIINRWQEVGKA	444		

QY	370	MYAPPIRGQIRCSSNITGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIE	429		
Db	445	MYALPIKGLIRCSSNITGLLLTRDGGGE-NQTTEIFRPGGDMRDNRSELYKYKVVKIE	503		
QY	430	PLGVAPTAKRRVVQREKRAV-TLGAMFLGLGAAGSTMGARSLTLTVQARQLLSGIVQQ	488		
Db	504	PLGVAPTAKRRVVQREKRAVGMGLGAMFLGLGAAGSTMGATSMALTVMQARQLLSGIVQQ	563		
QY	489	QNNLLRAIEAQHLLQLTVWGIK-LQARVLAVERYLKDQQLLGWGCSEGLICTTAVPWN	547		
Db	564	QNNLLRAIKAQHLLQLTVWGIKQLQARILAVERYLKDQQLLGWGCSEGLICTTAVPWN	623		
QY	548	ASWSNKSLDQIWNMTWMEWEREIDNYTNLIYTLIEESQNQOEKNEQELLELDKWASLWN	607		
Db	624	ASWSNKTLDQIWNMTWMEWDREIDNYTHLIYTLIEESQNQOEKNEQELLELDKWASLWT	683		
QY	608	WFDISKWLWYIK	619		
Db	684	WSDITKWLWYIK	695		
RESULT 15					
ENV_HV1H3					
ID_	ENV_HV1H3	STANDARD;	PRT;	856 AA.	
AC	P04624;				
DT	13-AUG-1987 (Rel. 05, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Envelope glycoprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].				
GN	ENV.				
OS	Human immunodeficiency virus type 1 (HXB3 isolate) (HIV-1).				
OC	Viruses; Retroviral viruses; Retroviridae; Lentivirus.				
OX	NCBI_TaxID=11707;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=85228248; PubMed=2988795;				
RA	Crowl R., Ganguly K., Gordon M., Conroy R., Schaber M., Kramer R., Shaw G.M., Wong-Staal F., Reddy E.P.;				
RA	"HTLV-III env gene products synthesized in E. coli are recognized by antibodies present in the sera of AIDS patients.";				
RT	Cell 41:979-986(1985).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M14100; AAA44679.1; --				
DR	PDB; 1JAU; 17-OCT-01.				
DR	PDB; 1JAV; 17-OCT-01.				
DR	HIV; M14100; ENV\$HXB3.				
DR	InterPro; IPR000328; Env GP41.				
DR	InterPro; IPR000777; GP120.				
DR	Pfam; PF00516; GP120; 1.				
DR	Pfam; PF00517; GP41; 1.				
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal; 3D-structure.				
KW	SIGNAL	1	30		
FT	CHAIN	31	511		EXTERIOR MEMBRANE GLYCOPROTEIN.
FT	CHAIN	512	856		TRANSMEMBRANE GLYCOPROTEIN.
FT	DISULFID	119	205		BY SIMILARITY.
FT	DISULFID	126	196		BY SIMILARITY.
FT	DISULFID	131	157		BY SIMILARITY.
FT	DISULFID	218	247		BY SIMILARITY.
FT	DISULFID	228	239		BY SIMILARITY.
FT	DISULFID	296	331		BY SIMILARITY.
FT	DISULFID	378	445		BY SIMILARITY.
FT	DISULFID	385	418		BY SIMILARITY.

FT	CARBOHYD	88	88	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	136	136	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	141	141	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	156	156	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	160	160	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	186	186	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	197	197	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	230	230	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	234	234	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	241	241	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	262	262	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	276	276	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	289	289	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	295	295	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	301	301	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	332	332	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	339	339	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	356	356	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	386	386	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	392	392	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	397	397	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	406	406	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	448	448	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	463	463	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	611	611	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	616	616	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	624	624	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	637	637	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	674	674	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	750	750	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	816	816	N-LINKED	(GLCNAC. . .)	(POTENTIAL) .
SQ	SEQUENCE	856 AA;	97188 MW;	3373C68BB84C1AFC	CRC64;	

Query Match: 84.4%; Score 2808; DB 1; Length 856;

Best Local Similarity 82.3%; Pred. No. 6.2e-210;

Matches 541; Conservative 33; Mismatches 43; Indels 40; Gaps 10;

Qy	1	SAVEKLVTVYVYGVVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIVLEN	60
Db	29	SATEKLVTVYVYGVVWKEATTTLFCASDAKAYDTEVHNVWATHAGVPTDPNPQEVVLVN	88
Qy	61	VTFNFMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKE -	119
Db	89	VTFNFMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSGRMI	148
Qy	120	MDRGEIKNCSFKVGA---GK-----LINCNTSVITQACPKV	152
Db	149	MEKGEIKNCSFNISTIRGKVQKEYAFFYKLDIIPIDNDTTSYTLTSCNTSVITQACPKV	208
Qy	153	SPEPIPIHYCAPAGFAILKCNDDKFGNGSGCTNVSTVQCTHGIRPVVSTQLLNGSLAE	212
Db	209	SPEPIPIHYCAPAGFAILKCNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE	268
Qy	213	GVWIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSITI--GPGRAFYATGDIIGDIR	270
Db	269	EVWIRSVNFTDNAKTIIVQLNTSVEINCTRPNNNTRKKIRIQRGPGRAPVTIGK-IGNMR	327
Qy	271	QAHCNISGEKNWNTLKQIVTKLQAQFG-NKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNS	329
Db	328	QAHCNISRAKNWATLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNS	387
Qy	330	TQLFNSTWNN----TIGPNNTNG--TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSS	383
Db	388	TQLFNSTWFNSTWSTEGSNNTGSDTITLPCRIKQFINMWQEVGKAMYAPPISGQIRCSS	447
Qy	384	NITGLLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV	443
Db	448	NITGLLLLTRDGGNN--NNGSEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVV	506
Qy	444	QREKRAVTLGAMFLGFLGAAGSTMGARSLLTVQARQLLSGIVQOQNNLLRAIEAQOHL	503
Db	507	QREKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQOQNNLLRAIEAQOHL	566

Qy	504	QLTVWGIK-LQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWNSKSLDQIWNMM	562
Db	567	QLTVWGIKQLQARILAVERYLKDQQLLGIWCGSGKLLCTTAVPWNASWNSKSLQIWNHT	626
Qy	563	TWMEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK	619
Db	627	TWMEWDREINNNTSLIHSLEIESQOQEKNEQELLELDKWASLWNWFNITNWLWYIK	683

Search completed: February 25, 2004, 14:17:01
Job time : 12.7652 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: February 25, 2004, 14:12:26 ; Search time 39.1462 Seconds
(without alignments)
4989.132 Million cell updates/sec

Title: US-09-891-609A-4
Perfect score: 3327
Sequence: 1 SAVEKLWTVVYGVVPWKEA.....DKWASLWNWFDISKWLWYIK 619

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25.*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3176	95.5	847	15 Q8AR21	human immun
2	2978.5	89.5	854	15 Q7SVL3	human immun
3	2971.5	89.3	850	15 Q7SVL4	human immun
4	2968.5	89.2	860	15 Q7SVL7	human immun
5	2967.5	89.2	852	15 Q7SVL5	human immun
6	2941	88.4	853	15 Q7SVL6	human immun
7	2925	87.9	843	15 Q70150	human immun
8	2924.5	87.9	854	15 O40222	human immun
9	2914.5	87.6	846	15 Q9PXW7	human immun
10	2906	87.3	847	15 Q75760	human immun
11	2900.5	87.2	852	15 O41883	human immun
12	2895.5	87.0	854	15 Q78225	human immun
13	2895	87.0	851	15 O56110	human immun
14	2894	87.0	854	15 O92762	human immun
15	2893	87.0	853	15 O56108	human immun
16	2892	86.9	853	15 Q03811	human immun

17	2890	86.9	853	15 Q7ZJF5	human immun
18	2889.5	86.9	852	15 O92761	human immun
19	2887.5	86.8	856	15 Q77694	human immun
20	2886.5	86.8	858	15 Q7SVL1	human immun
21	2886	86.7	845	15 Q70679	human immun
22	2883.5	86.7	850	15 Q74812	human immun
23	2883.5	86.7	860	15 Q7ZJD7	human immun
24	2882.5	86.6	864	15 Q8UL53	human immun
25	2881.5	86.6	856	15 Q72993	human immun
26	2878	86.5	859	15 Q7SUT0	human immun
27	2876.5	86.5	855	15 Q03805	human immun
28	2875	86.4	855	15 Q8Q367	human immun
29	2874.5	86.4	868	15 Q9WJW5	human immun
30	2874	86.4	859	15 Q7SVL2	human immun
31	2871	86.3	849	15 Q77368	human immun
32	2870.5	86.3	867	15 Q7ZC00	human immun
33	2869	86.2	861	15 Q9IUY9	human immun
34	2869	86.2	863	15 Q77989	human immun
35	2868	86.2	859	15 Q7SUS9	human immun
36	2867.5	86.2	852	15 O41885	human immun
37	2867	86.2	853	15 Q7SIK0	human immun
38	2867	86.2	861	15 Q9WJV5	human immun
39	2865	86.1	861	15 Q42031	human immun
40	2865	86.1	854	15 Q7SUQ7	human immun
41	2864.5	86.1	861	15 O56109	human immun
42	2864	86.1	846	15 Q8UL63	human immun
43	2863.5	86.1	853	15 Q7SUR3	human immun
44	2863	86.1	854	15 Q7SUQ9	human immun
45	2862.5	86.0			

ALIGNMENTS

RESULT 1

Q8AR21 Q8AR21 PRELIMINARY; PRT; 847 AA.
AC Q8AR21;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHIVSF162P3.2;
RA Gao F.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBBJ databases.
DR EMBL; AF536757; AAN05642.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
SQ SEQUENCE 847 AA; 95879 MW; 9CF863B97B3BB54C CRC64;

Query Match 95.5%; Score 3176; DB 15; Length 847;
Best Local Similarity 92.1%; Pred. No. 2e-250;
Matches 596; Conservative 7; Mismatches 16; Indels 28; Gaps 2;

Qy	1	SAVEKLWTVVYGVVPWKEATTTLFCASDAKAYDTEVHNWATHACVPTDNPQEI	VLEN 60
Db	28	SAVEKLWTVVYGVPAWKEATTTLFCASDAKAYDTEVHNWATHACVPTDNPQEI	VLEN 87
Qy	61	VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM	120
Db	88	VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTTSSNWKEM	147
Qy	121	DRGEIKNCSEFKVGAG-----KLINCNSTSVITQACPKVS	153

```
Db      :||||| |
148 NRGEIKNCSENVTTSIGNKMQKEYALFYRLDVVPIDNDNTSYNLINCNTSVITQACPKVS 207

Qy      :||||| |
154 FEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG 213
||||| |
208 FEPIPIHYCAPAGFAILKCNDKKFNGSGPCINVSTVQCTHGIRPVVSTQLLNGSLAEEG 267

Qy      :||||| |
214 VVIRSEFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273
||||| |
268 VVIRSEFTDNVKTIIIVQLKESVEINCTRPNNNTRKSIPIGPKAFYATGDIIGDIRQAH 327

Qy      :||||| |
274 CNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIIVMHSFNCGGEFFYCNSTQLF 333
||||| |
328 CNISGEKWNNTLKQIVTKLQAQFENKTIVFKQSSGGDPEIIVMHSFNCGGEFFYCNSTQLF 387

Qy      :||||| |
334 NSTWNNTIGPNNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIIRCSSNITGLLLTRD 393
||||| |
388 NSTWNNTIGPNNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIIRCSSNITGLLLTRD 447

Qy      :||||| |
394 GGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQREKRAVTLG 453
||||| |
448 GGREVGNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQREKRAVTLG 507

Qy      :||||| |
454 AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK-L 512
||||| |
508 AVFLGFLGAAGSTMGAASLTTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIKQL 567

Qy      :||||| |
513 QARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLLDQIWNNTWMEWEREID 572
||||| |
568 QARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLLDQIWNNTWMEWEREIG 627

Qy      :||||| |
573 NYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWFDISKWLWYIK 619
||||| |
628 NYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWLDISKWLWYIK 674
```

```
RESULT 2
Q7SVL3
ID Q7SVL3 PRELIMINARY; PRT; 854 AA.
AC Q7SVL3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81NV1;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247222; AAP37150.1; -.
KW Envelope protein.
SQ SEQUENCE 854 AA; 97048 MW; 36379231FF3AF12C CRC64;
```

Query Match 89.5%; Score 2978.5; DB 15; Length 854;
Best Local Similarity 85.9%; Pred. No. 2.8e-234;
Matches 562; Conservative 21; Mismatches 36; Indels 35; Gaps 5;

```
Qy      1 SAVEKLWVTYYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60
||| :|||||
28 SAAEQLWVTYYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEVVLEN 87

Qy      61 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTKSSNWKEM 120
|||||
28 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTTSSSWGTM 147
```

```
Qy      121 DRGEIKNCSFKVQAG-----KLLNCNTSVITQACPK 151
: |||||
148 ESGEIKNCSFNTTTSIRDKVQKEYALFYKLDVVPINDNTTTSYRLINCNTSVITQACPK 207

Qy      152 VSFEPPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 211
||| :|||||
208 VTFEPIPIHYCTPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 267

Qy      212 EGVVIRSEFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQ 271
||||| :|||||
268 EEVVRSDNFTDNAKTIIVQLNESVEINCTRPNNNTRKSIINIGPGRAFYATGEIIGDIRQ 327

Qy      272 AHCNISGEKWNNTLKQIVTKLQAQFG-NKTIVFKQSSGGDPEIIVMHSFNCGGEFFYCNST 330
||||| :|||||
328 AHCNLSRAKWNNTLKQIVTKLREQFGNNKTIIVFNQSSGGDPEIIVMHSFNCGGEFFYCDST 387

Qy      331 QLFNSTW---NNTIGPNNNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIIRCSSNITG 387
||||| ||| | :|
388 QLFNSTWENDNNIRGSNSTQENITLPCRIKQIINMWQGVGKAMYAPPPIRGQIIRCSSNITG 447

Qy      388 LLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQREK 447
||| :|||||
448 LLLTRDGGKDENGTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKARRVVQREK 507

Qy      448 RAV-TLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLT 506
||| :|||||
508 RAVGTIGAMFLGFLGAAGSTMGAASMALTVQARQLLSGIVQQNNLLRAIEAQHLLQLT 567

Qy      507 VWGIK-LQARVLAVERYLKDQQLLGIWCSGKLICTTAVPWNASWSNKSLLDQIWNNTWWM 565
|||||
568 VWGIKQLQARVLAVERYLKDQQLLGIWCSGKLICTTVPWNVSWSNKSLLDKIWNNTWWM 627

Qy      566 EWEREIDNYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWFDISKWLWYIK 619
|||||
628 EWEREIDNYTNLIYTLIEESONQOEKNEQELLELDKWASLWNWFSITNWLWYIK 681
```

```
RESULT 3
Q7SVL4
ID Q7SVL4 PRELIMINARY; PRT; 850 AA.
AC Q7SVL4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81NJ;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247221; AAP37149.1; -.
KW Envelope protein.
SQ SEQUENCE 850 AA; 96583 MW; 62ED5F6AB033D20B CRC64;
```

Query Match 89.3%; Score 2971.5; DB 15; Length 850;
Best Local Similarity 86.2%; Pred. No. 1e-233;
Matches 561; Conservative 24; Mismatches 33; Indels 33; Gaps 4;

```
Qy      1 SAVEKLWVTYYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEIIVLEN 60
||| :|||||
28 SAAEKLWVTYYGVVPVWKEATTLFCASDAKAYDTEVHNWVWATHACVPTDPNPQEVVLEN 87

Qy      61 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTKSSNWKEM 120
|||||
28 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTTSSSGGTM 147
```

```
QY 121 DRGEIKNCSPKVGAG-----KLINCNSTSVITQACPK 151
      :|||||:
Db 148 ERGEIKNCSENIITTSIRDKVQKEYALFYKLDVVPIDNDNTTTSYRLISCNSTSVITQACPK 207
      :|||||:
QY 152 VSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 211
      :|||||:
Db 208 VSFEPIPIHYCTPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 267
      :|||||:
QY 212 EGWIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQ 271
      :|||||:
Db 268 EEVWIRSDNFTDNAKTIIVQLKESVEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQ 327
      :|||||:
QY 272 AHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQ 331
      :|||||:
Db 328 AHCNLSRAKWDNTLKQIVRKLREQFGNKTIVFNQSSGGDPEIVTHSFNCGGEFFYCDSTQ 387
      :|||||:
QY 332 LFNSTWNNTIGPNTNG--TITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSNITGLL 389
      :|||||:
Db 388 LFNSTWNVTGSGNNTGNTITLPCRIKQIINMWQEVGKAMYAPPPIRGQIRCSNITGLL 447
      :|||||:
QY 390 LTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRA 449
      :|||||:
Db 448 LTRDGGNESE-TEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQREKRA 506
      :|||||:
QY 450 VTLAGMFLGFLGAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWG 509
      :|||||:
Db 507 VGIGAVFLGFLGAGSTMGAASMTLTVQARLLSGIVQQNNLLRAIEAQHLLQLTVWG 566
      :|||||:
QY 510 IK-LQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLDQIWNMTWMEWE 568
      :|||||:
Db 567 IKQLQARVLAVERYLKDQQLLGIWCGSGKLICTTVPWNASWSNKSLDKIWNMTWMEWE 626
      :|||||:
QY 569 REIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 619
      :|||||:
Db 627 REINNYTSLIYTLIEESNQOEKNEQELLELDKWASLWNWFDITKWLWYIK 677
      :|||||:
```

RESULT 4

```
Q7SVL7
ID Q7SVL7 PRELIMINARY; PRT; 860 AA.
AC Q7SVL7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81CA1;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247218; AAP37146.1; -.
KW Envelope protein.
SQ SEQUENCE 860 AA; 97762 MW; A7E185F2BD421590 CRC64;
```

```
Query Match      89.2%; Score 2968.5; DB 15; Length 860;
Best Local Similarity 84.4%; Pred. No. 1.9e-233;
Matches 557; Conservative 27; Mismatches 35; Indels 41; Gaps 3;

QY 1 SAVEKLVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60
      :|||||:
Db 28 SAAGKLVTVVYGVVPVWKEATTLFCASDAKAYDTEIHNWATHACVPTDPNPQEVALEN 87
      :|||||:
QY 61 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
      :|||||:
```

```
Db 88 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNATNTTSSSGGKM 147
      :|||||:
QY 121 DRGEIKNCSPKVGAG-----KLINCNSTSVITQACPK 151
      :|||||:
Db 148 EKGEIKNCSENIITNIRDQMKEYALFYRLDVVPIDNDSTNTSYRLISCNSTSVITQACPK 207
      :|||||:
QY 152 VSFEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 211
      :|||||:
Db 208 VSFEPIPIHYCTPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVISTQLLNGSLAE 267
      :|||||:
QY 212 EGWIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQ 271
      :|||||:
Db 268 EEVWIRSDNFTDNAKTIIVQLNESVEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQ 327
      :|||||:
QY 272 AHCNISGEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQ 331
      :|||||:
Db 328 AHCNLSRTKWNNTLRQIVYKLREQFGNKTIVFNQSSGGDPEIVTHSFNCGGEFFYCDSTQ 387
      :|||||:
QY 332 LFNSTWN--TIGPNTNGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIR 380
      :|||||:
Db 388 LFNSTWNNTLFNSTWNNDTKGSNNTEDITLPCRIKQIINMWQEVGKAMYAPPPIRGQIR 447
      :|||||:
QY 381 CSSNITGLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKKR 440
      :|||||:
Db 448 CSSNITGLLTRDGGNNKSETTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKKR 507
      :|||||:
QY 441 RVVQREKRAVTLGAMFLGFLGAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQ 500
      :|||||:
Db 508 RVVQREKRAVGIGAVFLGFLGAGSTMGAASMTLTVQARLLSGIVQQNNLLRAIEAQ 567
      :|||||:
QY 501 HLLQLTVWGIK-LQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLDQIW 559
      :|||||:
Db 568 HLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWCGSGKLICTTVPWNASWSNKSLDKIW 627
      :|||||:
QY 560 NMTWMEWEREIDNYTNLIYTLIEESNQOEKNEQELLELDKWASLWNWFDISKWLWYIK 619
      :|||||:
Db 628 NMTWMEWEREINNYTSLIYTLIEESNQOEKNEQELLELDKWASLWNWFDITKWLWYIK 687
      :|||||:
```

RESULT 5

```
Q7SVL5
ID Q7SVL5 PRELIMINARY; PRT; 852 AA.
AC Q7SVL5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81GA;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366(2003).
DR EMBL; AY247220; AAP37148.1; -.
KW Envelope protein.
SQ SEQUENCE 852 AA; 96680 MW; F8D811028025D998 CRC64;
```

```
Query Match      89.2%; Score 2967.5; DB 15; Length 852;
Best Local Similarity 85.6%; Pred. No. 2.2e-233;
Matches 559; Conservative 24; Mismatches 35; Indels 35; Gaps 4;

QY 1 SAVEKLVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60
      :|||||:
Db 28 SAAEQLVTVVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVGLEN 87
      :|||||:
QY 61 VTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120
      :|||||:
```


Db 88 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLGNTTNTSSGGGM 147
QY 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACPKVS 153
Db 148 ERGEIKNCSFNITTSIRDKVQKEYALLYKLDIVPIDDNNTSYRLISCNTSVITQACPKVS 207
QY 154 FEPIPIHYCAPAGFAILKCNCKFKNGSGPCNTVSTVQCTHGIRPVVSTQLLLNGSLAEEG 213
Db 208 FEPIPIHYCAPAGFAILKCRDKKFKNGKGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEE 267
QY 214 VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 273
Db 268 VVIRSDNFTDNAKTIIVQLKESVEINCTRPNNNTRKSIHIGPGRAFYTTGDIVGDIRQAH 327
QY 274 CNISGEKWNNTLKQIVTKLQAFGNKTIIVFKOSSGGDPEIVMHSFNCGGEFFYCNSTOLF 333
Db 328 CNISKAKWNTTLKQIVTKLREQFGNRTIVFNQSSGGDPEIVMHSFNCGGEFFYCNSTOLF 387
QY 334 N-----STWNNTIGPNNTNGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITG 387
Db 388 NSTWNASSTWNDEGSNNTGTTLPCRICKQIINMWQEVGKAMYAPPPIRGQIRCSSNITG 447
QY 388 LLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVWQREK 447
Db 448 LILTRDGGSN-ENDTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVWQREK 506
QY 448 RAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTV 507
Db 507 RAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLSGIVQQNNLLRAIEAQHLLQLTV 566
QY 508 WGIK-LQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLSLDIWNNTWME 566
Db 567 WGIKQLQARVLAMERYLKDQQLLGIWCGSGKLICTTVPWNASWSNKSLSLDIWNNTWME 626
QY 567 WEREIDNTNLIYTLIEESQNQKEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 627 WDREINNYTSLIYTLIEESQNQKEKNEQELLELDKWASLWNWFDITKWLWYIK 679

RESULT 6
Q7SVL6
ID Q7SVL6 PRELIMINARY; PRT; 853 AA.
AC Q7SVL6;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81CA2;
RX MEDLINE=22628496; PubMed=12743293;
RA Robbins K.E., Lemey P., Pybus O.G., Jaffe H.W., Youngpairoj A.S.,
RA Brown T.M., Salemi M., Vandamme A.M., Kalish M.L.;
RT "U.S. Human Immunodeficiency Virus Type 1 Epidemic: Date of Origin,
RT Population History, and Characterization of Early Strains.";
RL J. Virol. 77:6359-6366 (2003).
DR EMBL; AY247219; AAP37147.1; -.
KW Envelope protein.
SQ SEQUENCE 853 AA; 96545 MW; E3B2830A1261E237 CRC64;

Query Match 88.4%; Score 2941; DB 15; Length 853;
Best Local Similarity 84.8%; Pred. No. 3.2e-231;
Matches 556; Conservative 30; Mismatches 30; Indels 40; Gaps 5;

QY 1 SAVEKLWTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQIVLEN 60
Db 28 SATEKLWTVVYGVVPVWKEATTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLGN 87

QY 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNAATNTKSSNWKEM 120
Db 88 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRATNTTSSSGEKM 147
QY 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACPK 151
Db 148 EGGEIKNCSFNITTSRLDKMQKEYALPYKLDVVPIDNDNTNTSYRLISCNTSVITQACPK 207
QY 152 VSFEPIPIHYCAPAGFAILKCNCKFKNGSGPCNTVSTVQCTHGIRPVVSTQLLLNGSLAE 211
Db 208 VSFEPIPIHYCAPAGFAILKCNCKFKSGKGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAE 267
QY 212 EGVVIRSENFTDNAKTIIVOLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQ 271
Db 268 EEVVRSDNFTDNAKTIIVOLKEPVEINCTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQ 327
QY 272 AHCNISGEKWNNTLKQIVTKLQAFGNKTIIVFKOSSGGDPEIVMHSFNCGGEFFYCNSTQ 331
Db 328 AHCNLSITKWNNTLKQIVKLLKEQFGNKTIVFKOSSGGDPEIVTHSFNCGGEFFYCNSTK 387
QY 332 LFNSTWNN-----TIGPNNTNGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNIT 386
Db 388 LFNSTWNNSTWNGTEGSNNPEGNITLPCRICKQIINMWQEVGKAMYAPPPIRGQIRCSSNIT 447
QY 387 GLLLTRDGGKEISNT--TEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVWQ 444
Db 448 GLLLTRDGG--SNTSGTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVWQ 504
QY 445 REKRAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQ 504
Db 505 REKRAVGIGALFLGFLGAAGSTMGAASMTLVQARLLSGIVQQNNLLRAIEAQHLLQ 564
QY 505 LTVWGIK-LQARVLAVERYLKDQQLLGIWCGSGKLICTTAVPWNASWSNKSLSLDIWNNT 563
Db 565 LTVWGIKQLQARVLAVERYLKDQQLLGIWCGSGKLICTTVPWNASWSNKSLSLDIWNNT 624
QY 564 MMEWEREIDNTNLIYTLIEESQNQKEKNEQELLELDKWASLWNWFDISKWLWYIK 619
Db 625 MMEWEREINNYTSLIYTLIEESQNQKEKNEQELLELDKWASLWNWFDITKWLWYIK 680

RESULT 7
Q70150
ID Q70150 PRELIMINARY; PRT; 843 AA.
AC Q70150;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=014;
RX MEDLINE=95194694; PubMed=7888189;
RA Gao F., Yue L., Craig S., Thornton C.L., Robertson D.L.,
RA McCutchan F.E., Bradac J.A., Sharp P.M., Hahn B.H.;
RT "Genetic variation of HIV type 1 in four World Health Organization-
RT sponsored vaccine evaluation sites: generation of functional envelope
RT (glycoprotein 160) clones representative of sequence subtypes A, B, C,
RT and E. WHO Network for HIV Isolation and Characterization.";
RL AIDS Res. Hum. Retroviruses 10:1359-1368 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=014;
RX MEDLINE=95194691; PubMed=7545977;
RA WHO Global Programme on AIDS;
RT "HIV type 1 variation in World Health Organization-sponsored vaccine
RT evaluation sites: genetic screening, sequence analysis, and
RT preliminary biological characterization of selected viral strains. WHO
RT Network for HIV Isolation and Characterization.";

RL AIDS Res. Hum. Retroviruses 10:1327-1343(1994).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=014;

RX MEDLINE=96190564; PubMed=8627686;

RA Gao F., Morrison S.G., Robertson D.L., Thornton C.L., Craig S.,

RA Karlsson G., Sodroski J., Morgado M., Galvao-Castro B.,

RA von Briesen H., Beddows S., Weber J., Sharp P.M., Shaw G.M.,

RA Hahn B.H.;

RT "Molecular cloning and analysis of functional envelope genes from

RT human immunodeficiency virus type 1 sequence subtypes A through G. The

RT WHO and NIAID Networks for HIV Isolation and Characterization.";

RL J. Virol. 70:1651-1657(1996).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=014;

RA Allen E.E.;

RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.

DR EMBL; U08801; AAB05185.1; -.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR000328; Env GP41.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.

FT NON TER 1

SQ SEQUENCE 843 AA; 95679 MW; 14DB49A6D2F5FCE4 CRC64;

Query Match 87.9%; Score 2925; DB 15; Length 843;

Best Local Similarity 84.9%; Pred. No. 6.5e-230;

Matches 551; Conservative 25; Mismatches 39; Indels 34; Gaps 5;

QY 1 SAVEKLWVTVYGVVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60

DB 26 SAAEKLWVTVYGVVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLGN 85

QY 61 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

DB 86 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNATNTSST---M 141

QY 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACPKVSF 154

DB 142 EGGEIKNCSFNITTSIKTKVKDYALFYKLDVVPIDNDNTSYRLINCNTSVITQACPKVSF 201

QY 155 EPIPIHYCAPAGFAILKCNCKKFGNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEGV 214

DB 202 EPIPIHYCTPAGFAILQCNNKKFNGTGCPCTNVSTVQCTHGIRPVVSTQLLNGSLAEDEV 261

QY 215 VIRSENFDTNAKTIIVQLKESVEINCTRNPNNTKRSITIGPGRAFYATGDIIGDIRQAH 274

DB 262 VIRSSNFTDNARVLIIVQLNESVEINCTRNPNNTKRSIHLGPRAWYTTGTQIIGDIRQAH 321

QY 275 NISGEKWNNTLKQIVTKLQAOFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFN 334

DB 322 NLSSTKWNNTLRQITTEKLRQFGNKTIVFNQSSGGDPEIVMHSFNCGGEFFYCNSTQLFN 381

QY 335 STWNNTIGPNNT--NGTITLPCRIKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTR 392

DB 382 STWNTDSTWNNNTNGTITLPCRIKQIIVNMQEVGKAMYAPPPIRGQIRCSSNITGLLLTR 441

QY 393 DGGKEISNTTETFRPGGDMRDNRWRSLEYKYVVKVIEPLGVAPTAKRRVVQREKRAV-T 451

DB 442 DGGNSENKTTETFRPGGDMRDNRWRSLEYKYVVKVIEPLGVAPTAKRRVVQREKRAVGT 501

QY 452 LGAMFLGFLGAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK 511

DB 502 IGAMFLGFLGTAGSTMGAASITLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVWGIK 561

QY 512 -LQARVLAVERYLKDQQLGIWCGSKLICTTAVPWNASWSNKSLDQIWNMTWMEWERE 570

Db 562 QLQARVLAVERYLKDQQLGIWCGSKLICTTAVPWNASWSNKSLDKIWNMTWMEWERE 621

QY 571 IDNYTNLIYTLIEESQNOQEKNEQELLELDKASLWNNWFDISKWLWYIK 619

Db 622 IDNYTREIYTLIEESQNOQEKNEQELLELDKASLWNNWFDITKWLWYIK 670

RESULT 8

O40222

ID O40222 PRELIMINARY; PRT; 854 AA.

AC O40222;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Env polyprotein.

GN ENV.

OS Human immunodeficiency virus 1.

OC Viruses; Retroold viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AD8;

RX MEDLINE=96432129; PubMed=8835195;

RA Theodore T.S., Englund G., Buckler-White A., Buckler C.E.,

RA Martin M.A., Peden K.W.C.;

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF004394; AAB64170.1; -.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR000328; Env GP41.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.

SQ SEQUENCE 854 AA; 97291 MW; 06C45E9103C6C12 CRC64;

Query Match 87.9%; Score 2924.5; DB 15; Length 854;

Best Local Similarity 85.1%; Pred. No. 7.2e-230;

Matches 558; Conservative 17; Mismatches 42; Indels 39; Gaps 7;

QY 1 SAVEKLWVTVYGVVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIIVLEN 60

DB 28 SAVENLWVTVYGVVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLEN 87

QY 61 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

DB 88 VTENFNMMKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNVTNNINSS--EG 145

QY 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACPKVS 153

DB 146 MRGEIKNCSFNITTSIRDVKVKDYALFYRLDVVPIDNDNTSYRLINCNTSVITQACPKVS 205

QY 154 FEPIPIHYCAPAGFAILKCNCKKFGNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAE 213

DB 206 FEPIPIHYCTPAGFAILKCKDKKFGNGTGCPCKNVSTVQCTHGIRPVVSTQLLNGSLAE 265

QY 214 VVIRSENFDTNAKTIIVQLKESVEINCTRNPNNTKRSITIGPGRAFYATGDIIGDIRQAH 273

DB 266 VVIRSSNFTDNAKNIIVQLKESVEINCTRNPNNTKRSIHIGPGRAFYTTGDIIGDIRQAH 325

QY 274 CNISGEKWNNTLKQIVTKLQAOFG-NKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQL 332

Db	326	CNISRTKWNNTLNQIATKLKEQGNKTIIVFNQSSGGDPEIVMHSFNCGGEFFYCNSTQL	385
QY	333	FNSTWN-----NTIGPNNTNG--TITLPCRITKQIINRWQEVGKAMYAPPPIRGQIRCSSNI	385
Db	386	FNSTWNFNGTWNLTQSNGTGNDTITLPCRITKQIINMWQEVGKAMYAPPPIRGQIRCSSNI	445
QY	386	TGLLLTRDGGKEISNTTEIFRPGGGMRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQR	445
Db	446	TGLILTRDGGNNHNDTETFRPGGGMRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQR	505
QY	446	EKRAV-TLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQ	504
Db	506	EKRAVGTIGAMFLGFLGAAGSTMGAASITLTVQARLLLSGIVQQNNLLRAIEAQHLLQ	565
QY	505	LTVWGIK-LQARVLAVERYLKDQQLLGWCGSKLICTTAVPWNASWSNKSLDQIWNMT	563
Db	566	LTVWGIKQLQARVLAVERYLRDQQLLGWCGSKLICTTAVPWNASWSNKTLDMIWNMT	625
QY	564	WMEWEREIDNYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK	619
Db	626	WMEWEREIDNYTGLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDITNWLWYIK	681
RESULT 9			
Q9PXW7	ID	Q9PXW7	PRELIMINARY; PRT; 846 AA.
AC	Q9PXW7;		
DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)	
DE	GPI60,	envelope glycoprotein.	
GN	ENV.		
OS	Human immunodeficiency virus 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11676;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93189881; PubMed=8446773;		
RA	Schmidt-mayerova H., Gayet O., Guettari N., Bolmont C., Hirsch I.,		
RA	Chermann J.C.;		
RT	"Characterization of HIV1-PAR, a macrophage-tropic strain: cell		
RT	tropism, virus/cell entry and nucleotide sequence of the envelope		
RT	glycoprotein.";		
RL	Res. Virol. 144:21-26(1993).		
DR	GO; GO:0016021; C:integral to membrane; IEA.		
DR	GO; GO:0019028; C:viral capsid; IEA.		
DR	GO; GO:0019031; C:viral envelope; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	InterPro; IPR000328; Env GP41.		
DR	InterPro; IPR000777; GPI20.		
DR	Pfam; PF00516; GPI20; 1.		
DR	Pfam; PF00517; GP41; 1.		
KW	AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.		
SQ	SEQUENCE 846 AA; 96490 MW; 3D561D5735F90C84 CRC64;		
Query Match 87.6%; Score 2914.5; DB 15; Length 846;			
Best Local Similarity 84.7%; Pred. No. 4.7e-229;			
Matches 548; Conservative 27; Mismatches 43; Indels 29; Gaps 3;			
QY	1	SAVEKLVWTVVYGVVPVKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQIVLEN	60
Db	28	SATDKLVWTVVYGVVPVKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQIVLEN	87
QY	61	VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM	120
Db	88	VTEYFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLNATNTNSRNGMM	147
QY	121	DRGEIKNCSPKVGAG-----KLNCSNTSVITQACPVS	153
Db	148	EKGEIKNCSPNITTSIRDKVQKEYALFYKLDVVPIDNNSTRYRLISNTSVITQACPVT	207
QY	154	FEEPIHYCAPAGFAILKCNKFKNGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEEG	213
Db	208	FEEPIHYCAPAGFAILKCKKFDGKGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEE	267
QY	214	VVIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAPYATGDIIGDIRQAH	273
Db	268	VVIRSDNFTNAKTIIVQLNESVEIKTRPNNNTRKSIPIGPRAPYTTGEIIGDIRQAH	327
QY	274	CNISGEKWNNTLKQIVTKLQAQFGNKTIIVFKSSGGDPEIVMHSFNCGGEFFYCNSTQLF	333
Db	328	CTISKTKWENTFKQIVRKLREQYKNKTIIVFNQSSGGDPEIVTHSFNCGGGEFFYCNSTQLF	387
QY	334	NSTWNNTIGPNNTNGTITLPCRITKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGLLLTRD	393
Db	388	NSTWNTGESNDTERTITLPCRITKQIINMWQKVGMKYAPPPIRGQIRCSSNITGLLLTRD	447
QY	394	GGKEISNTTEIFRPGGGMRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVTLG	453
Db	448	GGNNNNNGTEIFRPVGGMRDNWRSELYKYKVVKIEPLGVAPTAKRRVVQREKRAVGLG	507
QY	454	AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQTLVWGIK-L	512
Db	508	AMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQTLVWGIKQL	567
QY	513	QARVLAVERYLKDQQLLGWCGSKLICTTAVPWNASWSNKSLDQIWNNTWMEWEREID	572
Db	568	QARVLAVERYLKDQILGILGWSGSKSICTTAVPWNASWSNKSLD-IWNNTWMEWEREID	626
QY	573	NYTNLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISKWLWYIK	619
Db	627	NYTNLIYTLIEESQNQQEKNDDELLELDKWASLWNWFDITRWLWYIK	673
RESULT 10			
Q75760	ID	Q75760	PRELIMINARY; PRT; 847 AA.
AC	Q75760;		
DT	01-NOV-1996	(TrEMBLrel. 01, Created)	
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Envelope	glycoprotein.	
GN	ENV.		
OS	Human immunodeficiency virus 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11676;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=JRFL;		
RX	MEDLINE=87206194; PubMed=3646751;		
RA	Koyanagi Y., Miles S., Mitsuyasu R.T., Merrill J.E., Vinters H.V.,		
RA	Chen I.S.;		
RT	"Dual infection of the central nervous system by AIDS viruses with		
RT	distinct cellular tropisms.";		
RL	Science 236:819-822(1987).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=JRFL;		
RX	MEDLINE=91043044; PubMed=2172833;		
RA	O'Brien W.A., Koyanagi Y., Namazie A., Zhao J.Q., Diagne A., Idler K.,		
RA	Zack J.A., Chen I.S.;		
RT	"HIV-1 tropism for mononuclear phagocytes can be determined by regions		
RT	of gp120 outside the CD4-binding domain.";		
RL	Nature 348:69-73(1990).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=JRFL;		
RX	MEDLINE=92092169; PubMed=1684385;		
RA	Pang S., Vinters H.V., Akashi T., O'Brien W.A., Chen I.S.;		
RT	"HIV-1 env sequence variation in brain tissue of patients with AIDS-		
RT	related neurologic disease.";		
RL	J. Acquir. Immune Defic. Syndr. 4:1082-1092(1991).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=JRFL;		
RA	Pang S., Vinters H.V., Akashi T., O'Brien W.A., Chen I.S.,		

Db 625 MEWEREIDNYTSLIYSLIEESQOQVKNEQELLELDKWASLWSWFDITKWLWYIK 679

RESULT 12

Q78225 PRELIMINARY; PRT; 854 AA.

AC Q78225;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Env protein.

GN ENV.

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Ba-1;

RX MEDLINE=91289160; PubMed=1905842;

RA Hwang S.S., Boyle T.J., Lyerly H.K., Cullen B.R.;

RT "Identification of the envelope V3 loop as the primary determinant of

RT cell tropism in HIV-1.";

RL Science 253:71-74(1991).

DR EMBL; M63929; AAA75116.1; -.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR000328; Env GP41.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.

SQ SEQUENCE 854 AA; 97186 MW; 005FC881B5CBDFD8 CRC64;

Query Match 87.0%; Score 2895.5; DB 15; Length 854;

Best Local Similarity 83.4%; Pred. No. 1.7e-227;

Matches 546; Conservative 36; Mismatches 36; Indels 37; Gaps 5;

QY 1 SAVEKLVWTVVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVLEN 60

Db 28 NAEKLVWTVVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVLEN 87

QY 61 VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNA-----TNTKSS 115

Db 88 VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNATNGDNTTSS 147

QY 116 NWKEMDRGEIKNCSFKVGAG-----KLINCNTSVITQ 147

Db 148 SREMMGGGEMKNCSEFKITTIRGKVQKEYALFYELDIVPIDNNSNNRYRLISCNTSVITQ 207

QY 148 ACPKVSFEPIPIHYCAPAGFAILKCNCKKFNCGSPCTNVSTVQCTHGIRPVVSTQLLLNG 207

Db 208 ACPKISFEPIPIHYCAPAGFAILKCKDKKFNKGKPCSNVSTVQCTHGIRPVVSTQLLLNG 267

QY 208 SLAEFGVIRSENFNTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIG 267

Db 268 SLAEFVVIRSENFADNAKTIIVQLNESVEINCTRPNNNTRKSIHIGPGRALYTTGEIIG 327

QY 268 DIRQAHCNISGEKNNTLKQIVTKLQAFQGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYC 327

Db 328 DIRQAHCNLSRAKWNDTLNKIVIKLREQFGNKTIVFKHSSGGDPEIVTHSFNCGGEFFYC 387

QY 328 NSTQLNSTWNNTTIGPNT--NGTITLPCRICKQIINRWQEVGKAMYAPPIRGQIRCSSNI 385

Db 388 NSTQLNSTWNVTESNNTVENNTITLPCRICKQIINMWQKVGKAMYAPPIRGQIRCSSNI 447

QY 386 TGLLLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQR 445

Db 448 TGLLLTRDGGPE-ANKTEVFRPGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRVVQR 506

QY 446 EKRAVTLGAMFLGFLGAAGSTMGARSLLTQVQARQLLSGIVQOQNNLLRAIEAQHLLQL 505

Db 507 EKRAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQOQNNLLRAIEAQHLLQL 566

QY 506 TVWGIK-LQARVLAVERYLKDOQLGIWCGSKLICTTAVPWNASWSNKSLDQIWNMTW 564

Db 567 TVWGIKQLQARVLAVERYLKDOQLGIWCGSKLICTTAVPWNASWSNKSINKIWDNMTW 626

QY 565 MEWEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 627 MEWDREINNYSIIYSLIEESQOQEKNEQELLELDKWASLWNWFDITEWLWYIK 681

RESULT 13

O56110 PRELIMINARY; PRT; 851 AA.

AC O56110;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Envelope glycoprotein.

GN ENV.

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SFMHS8.1;

RX MEDLINE=98178716; PubMed=9519894;

RA McCutchan F.E., Sanders-Buell E., Salminen M.O., Carr J.K.,

RA Sheppard W.H.;

RT "Diversity of the human immunodeficiency virus type 1 envelope

RT glycoprotein in San Francisco Men's Health Study participants.";

RL AIDS Res. Hum. Retroviruses 14:329-337(1998).

DR EMBL; AF025756; AAC40593.1; -.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0006096; P:glycolysis; IEA.

DR InterPro; IPR000328; Env GP41.

DR InterPro; IPR000173; GAP_dhdrogenase.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

DR PROSITE; PS00071; GAPDH; 1.

KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.

SQ SEQUENCE 851 AA; 96516 MW; 1CDAC83EE6464531 CRC64;

Query Match 87.0%; Score 2895; DB 15; Length 851;

Best Local Similarity 83.1%; Pred. No. 1.8e-227;

Matches 542; Conservative 36; Mismatches 40; Indels 34; Gaps 4;

QY 1 SAVEKLVWTVVYGVVPVWKEATTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVLEN 60

Db 28 SAADKLVWTVVYGVVPVWKEATTLFCASDAKAYETEVEHNVWATHACVPTDPNPQEVLEN 87

QY 61 VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM 120

Db 88 VTENFNWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDLRNDTNTSSSGETM 147

QY 121 DRGEIKNCSFKVGAG-----KLINCNTSVITQACPKVSF 154

Db 148 ERGEIKNCSFNITTSIRDVKVQKEYALLHKLDDVVPIDNTSYRLVSCNTSVITQACPKVSF 207

QY 155 EPIPIHYCAPAGFAILKCNCKKFNCGSPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEGV 214

Db 208 EPIPIHYCAPAGFAILKCNCKRFGNGKGPCTNVSTVQCTHGIRPVVSTQPLLNGSLAEKEV 267

QY 215 VIRSENFNTDNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQAH 274

Db 268 VIRSDNFTDNAKTIIVQLNESVEINCIIRPNNNTRKSIINIGPGRAFYTTGEIIGDIRQAH 327

QY 275 NISGEKNNTLQKIVTKLQAFQGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLF- 333

Db 328 NLSEAKWNHTLEQIAKKLREQFGNKTIVFNQSSGGDPEIVMYSFNCGGEFFYCNSTKLFN 387

Qy 334 -----NSTWNNTIGPNNTNGTITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGL 388

Db 388 RTWSVNSTWNTDEGVNNTGENITLPCRICKQIINMWQEVGKAMYAPPPIKGQIRCSSNITGL 447

Qy 389 LLTRDGGKEISNTTEIFRPGGDMRDNRSELYKYVVKIEPLGVAPTAKRRVVQREKR 448

Db 448 LLTRDGGNS-NNETETFRPGGDMRDNRSELYKYVVKIEPLGVAPTAKRRVVQREKR 506

Qy 449 AVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTVW 508

Db 507 AVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQNNLLRAIEAQHLLQLTVW 566

Qy 509 GIK-LOARVLAVERYLKDQQLLGWCGSGKLICTTAVPWNASWSNKSLDQIWNNTWMEW 567

Db 567 GIKQLQARVLAVERYLKDQQLLGWCGSGKLICTTVPWKTWSNKSLEKIWNNTWMEW 626

Qy 568 EREIDNNTLIYTLIEESQNQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 627 DREINNYTSLIYTLIEESQNQEKNEQELLELDKWASLWNWFDITNWLWYIK 678

RESULT 14

O92762

ID O92762 PRELIMINARY; PRT; 854 AA.

AC O92762;

DT 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)

DE Envelope glycoprotein.

GN ENV.

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SFMS6.3;

RX MEDLINE=98178716; PubMed=9519894;

RA McCutchan F.E., Sanders-Buell E., Salminen M.O., Carr J.K.,

RA Sheppard W.H.;

RT "Diversity of the human immunodeficiency virus type 1 envelope glycoprotein in San Francisco Men's Health Study participants.";

RL AIDS Res. Hum. Retroviruses 14:329-337(1998).

DR EMBL; AF025754; AAC40591.1; --.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR000328; Env_GP41.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.

SQ SEQUENCE 854 AA; 97287 MW; AFD8A82B62C5B27D CRC64;

Query Match 87.0%; Score 2894; DB 15; Length 854;

Best Local Similarity 83.6%; Pred. No. 2.2e-227;

Matches 546; Conservative 29; Mismatches 44; Indels 34; Gaps 6;

Qy 1 SAVEKLWTVTVYGVVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQIEIVLEN 60

Db 29 SATEKLWTVTVYGVVPVWKEATTTLFCASDAKAYDTEAHNIWATHACVPTDPNPQIEVDLEN 88

Qy 61 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTN-LKNATNTKSSNWKE 119

Db 89 VTENFNMWKNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLKCTNDLKNNTTNTNNSWEK 148

Qy 120 MDRGEIKNCSFKVGAG-----KLINCNTSVITQACPKV 152

Db 149 METGEIKNCSFNITNRRGKMQKEYALFYKLDVVSIDNDNTSYRLISCNSTSVITQACPKV 208

Qy 153 SFEPPIHYCAPAGFAILKCNDKKENGSGPCTNVSTVQCTHGIRPVVSTQLLNGSLAEE 212

Db 209 SFQPIPIHYCAPAGFAILKCCKKFSKGKGCINVTQCTHGIRPVVSTQLLNGSLAEE 268

Qy 213 GVVIRSENFDTNAKTIIVQLKESVEINCTRPNNNTRKSIITIGPGRAFYATGDIIGDIRQA 272

Db 269 EVVIRSDNFTDNAKTIIVQLKEPVEINCTRPNNNTRKSIPIGPGRAFYTTEIGIIGNIROA 328

Qy 273 HCNISGEKWNNTLKQIVTKLQAFGNKTIIVFKOSSGGDPEIVMHSFNCGGEFFYCNSTQL 332

Db 329 HCNISRAKWNNTLQQIIVIELKEHFKNKTIIVFNHSSGGDPEIVMHSFNCGGEFFYCNSTQL 388

Qy 333 FNSTWNNT--IGPNNTNG--TITLPCRICKQIINRWQEVGKAMYAPPPIRGQIRCSSNITGL 388

Db 389 FSSTWNTSESEGVNTEGNDTIILLCRICKQIIVNMWQEVGKAMYAPPPIRGQIRCSSNITGL 448

Qy 389 LLTRDGGKEI-SNTTEIFRPGGDMRDNRSELYKYVVKIEPLGVAPTAKRRVVQREK 447

Db 449 LLTRDGGNSNGSKAEVFRPGGDMRDNRSELYKYKIVKIEPLEVAPTAKRRVVQREK 508

Qy 448 RAVTLGAMFLGFLGAAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTV 507

Db 509 RAAGLGVMFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQNNLLRAIEAQHLLQLTV 568

Qy 508 WGIK-LOARVLAVERYLKDQQLLGWCGSGKLICTTAVPWNASWSNKSLDQIWNNTWME 566

Db 569 WGIKQLQARVLAVERYLKDQQLLGWCGSGKLICTTAVPNTWSNKSLDRIWNNTWME 628

Qy 567 WEREIDNNTLIYTLIEESQNQEKNEQELLELDKWASLWNWFDISKWLWYIK 619

Db 629 WEKEIDNNTSLIYTLIEESQNQEKNEQELLELDKWASSWNWFDITQWLWYIK 681

RESULT 15

O56108

ID O56108 PRELIMINARY; PRT; 853 AA.

AC O56108;

DT 01-JUN-1998 (TReMBLrel. 06, Created)

DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)

DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)

DE Envelope glycoprotein.

GN ENV.

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SFMS3.1;

RX MEDLINE=98178716; PubMed=9519894;

RA McCutchan F.E., Sanders-Buell E., Salminen M.O., Carr J.K.,

RA Sheppard W.H.;

RT "Diversity of the human immunodeficiency virus type 1 envelope glycoprotein in San Francisco Men's Health Study participants.";

RL AIDS Res. Hum. Retroviruses 14:329-337(1998).

DR EMBL; AF025751; AAC40589.1; --.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR000328; Env_GP41.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.

SQ SEQUENCE 853 AA; 96830 MW; EC8BA4418F8AEC47 CRC64;

Query Match 87.0%; Score 2893; DB 15; Length 853;

Best Local Similarity 84.2%; Pred. No. 2.7e-227;

Matches 550; Conservative 24; Mismatches 45; Indels 34; Gaps 4;

Qy 1 SAVEKLWTVTVYGVVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQIEIVLEN 60

Db 28 SAAEKLWTVTVYGVVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQIEVGLEN 87

Qy	61	VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLHCT-NLQNA TNTKSSNWKE	119
Db	88	VTENFNMWKNMVEQMHEDIISLWDSLKPCVKLTPLCVTLNRD DDLKNA TN TNSSGEM	147
Qy	120	MDRGEIKNC SFKVGAG-----K LINCNTSVITQACP	150
Db	148	EMRGEIKNC SFNITTSIRD KVKEYALFYKLDVVPINNNTTTSYRLINCNTSVITQACP	207
Qy	151	KVSFEPIPIHYCAPAGFAILKNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQ LLLNGSLA	210
Db	208	KVSFEPIPIHYCTPAGFAILKCKDKFNGTGPCTSVSTVQCTHGIRPVVSTQ LLLNGSLA	267
Qy	211	EEGWIRSENF TDNAKTIIVQLKESVEINCTRPNNTRKSIITIGPGRAFYATGDIIGDIR	270
Db	268	EEEVWIRSENF TDNAKIIIVQLNESVESNCTRPNNTRKSIPIGPGRAFYTTGEIIGEIR	327
Qy	271	QAHCNISGEKWNNTLKQIVTKLOAQFGNKTIIVFKQSSGGDPEIVMHSFNCGGEFFYCNST	330
Db	328	QAHCNLSRTKWDNALQQA KKLKGQFGTKTIVFNQSSGGDPEIVMHTFNCGGEFFYCNST	387
Qy	331	QLFNSTWNNTIGPNNNT---NGTITLPCRIKQIINRWQEVGKAMYAPP I RQIRCSSNITG	387
Db	388	QLFNSTWNVTRESNNTSTEGNGTITLPCRIKQIINRWQGVGKAMYAPP I RQIRCSSNITG	447
Qy	388	LLLTRDGGKEISNTTEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRRVVQREK	447
Db	448	LLLTRDGGNNESNETETFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRRVVQREK	507
Qy	448	RAVTLGAMFLGFLGAGSTMGARSLTLTVQARQLLSGIVQQNNLLRAIEAQOHLQLTV	507
Db	508	RAVGIGAVFLGFLGAGSTMGAASMTLTVQARLLLSGIVQQNNLLRAIEAQOHLQLTV	567
Qy	508	WGIK-LQARVLAVERYLKDQQLGIWCGSGKLICTTAVPWNASWSNKS L DQIWNNTWME	566
Db	568	WGIKQLQARVLAVERYLKDQQLGIWCGSGKLICTTVPWNASWSNKS LDKIWHNNTWME	627
Qy	567	WEREIDNYTNLIYTLIEESQOQEKNEQELLELDKWASLWNWFDISKWLWYIK	619
Db	628	WEREIDNYTSLIYTLIEESQOQEKNELELELDKWASLWNWFDITKWLWYIK	680